"EVEN STILZITTEN DAN",
BROMDEN DE VLIEGEN.
EN EVEN LATER:
"KAN NIET MISSEN
- DAT IS IBD"

DIAGNOSTIC STRATEGIES

IN CHILDREN WITH CHRONIC GASTROINTESTINAL SYMPTOMS

IN PRIMARY CARE

Gea Holtman

The study described in this thesis was funded by a grant from the Fonds NutsOhra and a grant from the University Medical Centre Groningen doelmatigheidsonderzoek.

Publication of this thesis was financially supported by Research Institute SHARE (School of HeAlth REsearch), University of Groningen, University Medical Centre Groningen, and Stichting Kleine Kwalen in Huisartsgeneeskunde.

Diagnostic strategies in children with chronic gastrointestinal symptoms in primary care

ISBN 978-90-367-9057-4 (printed version)
ISBN 978-90-367-9056-7 (electronic version)
Graphic design Teuntje van de Wouw
Printed by GVO drukkers & vormgevers B.V.

rijksuniversiteit groningen

Diagnostic strategies in children with chronic gastrointestinal symptoms in primary care

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

maandag 31 oktober 2016 om 14.30 uur

door

Geeske Atje Holtman

geboren op 10 september 1985
te Groningen

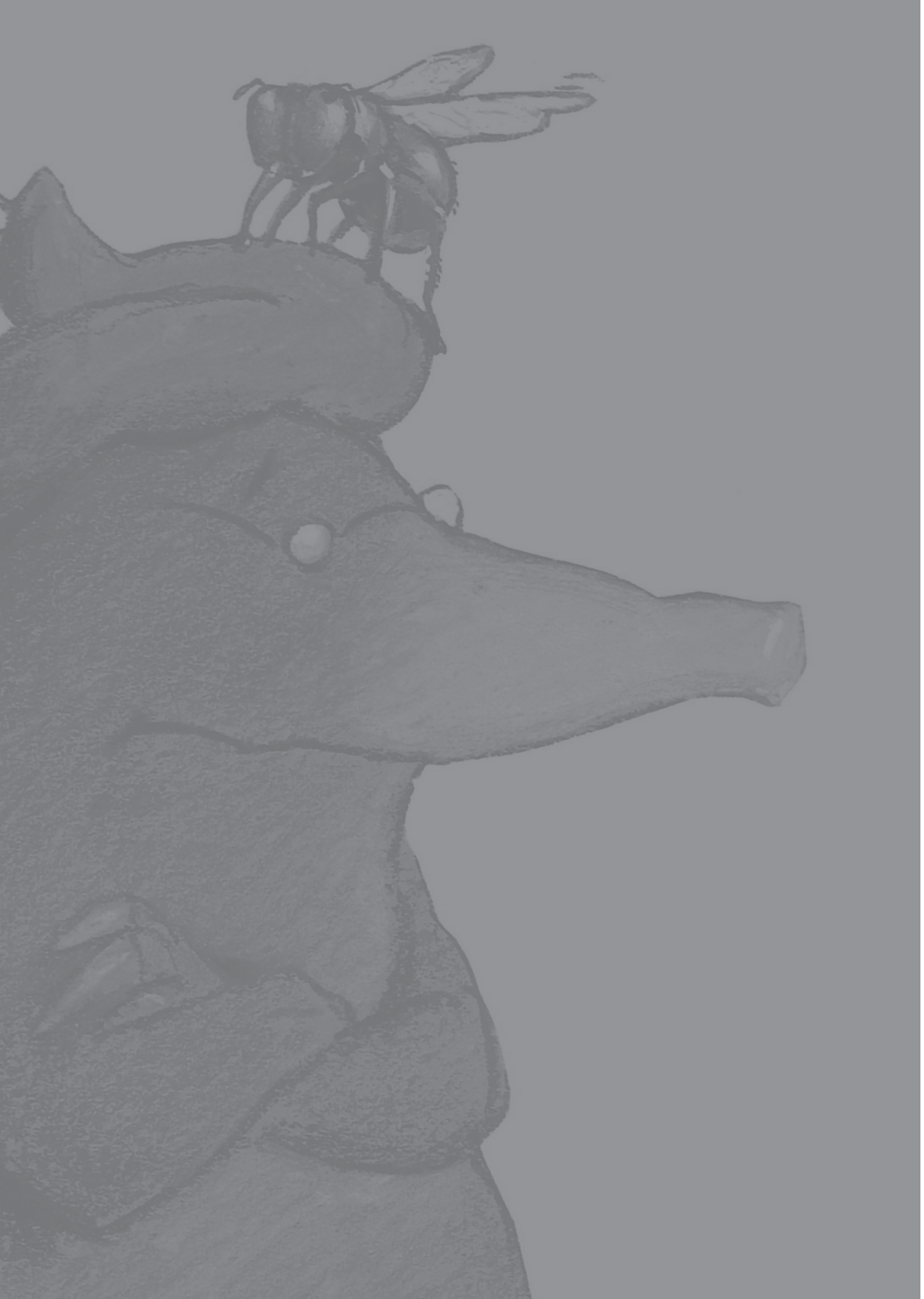
Promotor
Prof. dr. M.Y. Berger

Copromotores
Dr. Y. Lisman-van Leeuwen
Dr. P.F. van Rheenen

Beoordelingscommissie
Prof. dr. G.J. Dinant
Prof. dr. H.J. Verkade
Prof. dr. P.M.M. Bossuyt

In memory of my mother

CONTENTS



CHAPTER 1

GENERAL INTRODUCTION

Chronic or recurrent gastrointestinal symptoms are common presentations among children in primary care, where they account for approximately 2% to 5% of all childhood consultations.[1-3] In most cases, these symptoms are attributed to functional gastrointestinal disorders, which are characterized by recurrent or continuous abdominal pain without signs of any inflammatory, anatomic, metabolic, or neoplastic pathology. To avoid delay in diagnosis and treatment of possible organic disease, a thorough assessment of the differential diagnosis is necessary in these children. At the same time, it is important to avoid unnecessary referrals and extensive testing of children with functional gastrointestinal disorders, because this may progress to chronic abdominal pain without reassurance for either children or parents.[4,5] Indeed, failure to identify functional gastrointestinal disorders can have a major impact on both healthcare systems and the wellbeing of children.[6,7] Early recognition of, and appropriate coping strategies for, functional abdominal symptoms can enable faster recovery and prevent associated familial, psychological, and co-morbid conditions.[8,9] Because symptoms of functional gastrointestinal disorders may be indistinguishable from organic disease, it is a diagnostic challenge for physicians to differentiate between them accurately.

INFLAMMATORY BOWEL DISEASE

The differential diagnosis of chronic abdominal pain is broad and includes constipation, functional gastrointestinal disorders, gastroenteritis by parasites or colonic pathogens, celiac disease, and inflammatory bowel disease (IBD).[10] Of these, IBD is an organic disease, comprising Crohn's disease, ulcerative colitis, and IBD unclassified, for which general practitioners should be vigilant. Although the incidence of IBD is increasing among children aged younger than 18 years, the overall incidence remains low at 5.2/100,000 new cases per year in the Netherlands.[11,12] The clinical presentation of IBD can be very diverse and atypical.[13-16] While the combination of rectal bleeding and diarrhoea is the most common presentation of ulcerative colitis, Crohn's disease may present with abdominal pain, diarrhoea, anaemia, unexplained weight loss, or growth failure.[17] In children with suspected IBD, diagnosis should be made by endoscopy (ileocolonoscopy and oesophagogastroduodenoscopy) with biopsies.[18] However, these are invasive and expensive tests that are limited to specialist care facilities, requiring general anaesthesia or deep sedation in children.[18] It is also essential that appropriate therapy be started for IBD to reduce the inflammation (induction therapy), maintain remission (maintenance therapy), and improve the nutritional status and quality of life of the child.[19] Any delay in diagnosis and appropriate treatment for IBD may lead to complications such as anaemia, irreversible growth failure, and delayed sexual maturation.[20] In addition, children with IBD may experience social or emotional problems related to unrecognized IBD. Each of these problems can negatively influence the quality of life of children with IBD.[21,22] Therefore, early recognition and timely treatment is essential, and any assessment of chronic gastrointestinal symptoms would ideally be able to differentiate IBD from other potential causes of the symptoms.

SYMPTOMS, SIGNS, AND DIAGNOSTIC TESTS

In the diagnostic process of children with chronic gastrointestinal symptoms, symptoms and signs should be able to help clinicians safely exclude IBD in children with chronic

gastrointestinal symptoms, and help them select children who need further diagnostic assessment. However, common "alarm symptoms" (i.e., involuntary weight loss, rectal blood loss, family history of IBD, growth failure, extra-intestinal symptoms, and peri-anal lesions) assessed by history and physical examination, discriminate poorly between functional gastrointestinal disorders and IBD.[23,24] This diagnostic uncertainty, coupled with the difficult trade-off between avoiding unnecessary referral for invasive testing and not missing a severe case of chronic disease, means that a simple, readily available, and accurate test is needed. According to national and international guidelines, the general practitioner should refer children with chronic diarrhoea and/or recurrent abdominal pain for further diagnostic assessment when alarm symptoms or deviant blood marker results (i.e., c-reactive protein, erythrocyte sedimentation rate, haemoglobin, and platelet count) are present.[10,18,19] However, these commonly available blood markers are not specific for intestinal inflammation.[25,26] Moreover, the recommendations are based on test characteristics assessed in highly selected populations, and these test characteristics may vary across different settings.[27] To date, the added value of blood markers on alarm symptoms is unknown in symptomatic children.

There is no comprehensive overview, in either primary or specialist care, of readily available tests for the triage of children who may need further diagnostic assessment for IBD. Triage instruments are simple and inexpensive tests that can be used to determine which patients should receive the more invasive and expensive existing test. The triage tests may be less accurate that the existing tests and do not need to replace them.[28] An overview of triage tests could assist in suggesting which tests can exclude IBD safely, and which tests can best identify children who need further investigation. This information could improve the clinical decision making of doctors who encounter children with chronic gastrointestinal symptoms.

FAECAL CALPROTECTIN

Inflammation of the mucosal layer of the colon increases the excretion of neutrophil granulocytes into the bowel lumen. Calprotectin, first described in 1980, is a calcium-binding protein released from activated granulocytes, particularly neutrophils in plasma, tissue, and faeces.[29] Calprotectin resists enzymatic degradation and is stable in stool samples for up to seven days at room temperature, so faecal calprotectin levels can be used as a non-invasive diagnostic test for intestinal inflammation. Thus, a method was developed in 1992 to measure faecal calprotectin, and this was later improved in 2000. The improved method only required 50–100 mg of faeces instead of the 5 g required by the original test,[30] and allowed for stool samples to be collect at home and send to laboratories. Although this conventional testing of faecal calprotectin required laboratory facilities for enzyme-linked immunosorbent assay, point-of-care tests have now become available that allow the patient to bring a stool sample to the clinic and get a test result within 15 minutes.

There are other salient factors to consider with faecal calprotectin. Several studies have shown that calprotectin concentrations are increased in the faeces of adults with IBD, colon cancer, and colonic pathogens, but not in patients with functional gastrointestinal disorders[31,32] or coeliac disease.[33] Importantly, the calprotectin level has also been shown to be increased in children with IBD,[34,35] while results for coeliac and other organic diseases are limited in this population.[36] Children with functional gastrointestinal disorders had normal

or slightly elevated calprotectin levels.[37,38] Finally, although the recognized threshold is >50 µg/g faeces for intestinal inflammation, this threshold is not applicable to children younger than 4 years, because young healthy children also have high calprotectin concentrations in their faeces.[39]

EVIDENCE OF FAECAL CALPROTECTIN IN CHILDREN

Several meta-analyses have shown that faecal calprotectin has a very high sensitivity (0.92–0.98) and modest specificity (0.68–0.76) for IBD in referred children.[40-42] Therefore, this is an excellent test for excluding IBD when the result is normal, and can prevent unnecessary invasive endoscopies. Although these meta-analyses included studies performed in children with symptoms suggestive of IBD, all of the included studies were performed in children who had a high pre-test probability of disease. Indeed, the referred populations in these studies had a IBD prevalence of approximately 60%. Prevalence may reflect other mechanisms that influence the sensitivity and specificity, such as patient spectrum, referral filter, and reader expectations. These mechanisms may cause that sensitivity and specificity vary in different clinical populations.[43] The prevalence of IBD is much lower among non-referred children presenting in primary care, so it is reasonable to assume that this might influence the sensitivity and specificity of the tests in those settings. The results of the meta-analyses may, therefore, not be generalizable to children presenting in primary care. Before testing with faecal calprotectin can be recommended for children with gastrointestinal symptoms in primary care, more information is needed about the diagnostic accuracy of faecal calprotectin in this population.

CLINICAL RELEVANCE

A test is considered clinically relevant to primary care when it is useful, simple, and non-invasive, with good test characteristics; thus, a test should have a low false-negative rate (high sensitivity) and a low false-positive rate (high specificity).[44] Whether false-negative and false-positive rates are acceptable for patients and for decision making are dependent on the pre-test probability and the trade-off between false negatives and false positives. In children presenting with chronic gastrointestinal symptoms, the probability of having IBD is very low. Therefore, although failure to diagnose IBD is serious in these children, the impact of false positives is much higher than that of false negatives. These false positives may have serious consequences, including excessive testing and decreased wellbeing for children with functional gastrointestinal disorders. A low false-positive rate is important for any test assessing IBD in children presenting with chronic gastrointestinal symptoms in primary care. In contrast, the pre-test probability of IBD is much higher in children presenting with chronic gastrointestinal symptoms and alarm symptoms, and in this population the impact of false negatives becomes higher because of the greater likelihood of IBD. Therefore, false negatives are less acceptable than false positives, and the risks of referral of false positives are reduced, because children might have other organic diseases (e.g., celiac disease).

If the test characteristics are suitable to the primary care setting, it is important to evaluate the added value of testing to symptoms and signs measured by history and physical examination, which are a routine part of daily practice. A key question is whether additional

testing for faecal calprotectin can improve the value of the diagnostic assessment beyond that are already available.[45] If a test could be showed to have good test characteristics and added value to symptoms and signs, it may not only improving the clinical management of patients in primary care but also optimizing the referral of patients to specialist care.

METHODOLOGICAL RELEVANCE

The quality of the diagnostic process is essential for the quality of patient care.[46] The number of diagnostic techniques is progressively increasing today. However, the diagnostic value of tests is often not properly evaluated before use in practice, because diagnostic accuracy studies are not a prerequisite for being allowed on the market according to the European CE approval system. To determine the diagnostic accuracy of a test is an essential step in the evaluation of tests. The diagnostic accuracy of a test is evaluated by comparing the results of the test under evaluation with the results of the reference standard in the same patients. Ideally, a cohort of consecutive patients suspected for the condition of interest is included in which both the test(s) under evaluation and the reference standard is performed in all patients.

In primary care, there are two main methodological challenges in diagnostic accuracy studies. Firstly, the prior probability of the disease in patients with symptoms is often low in primary care. Studies performed in patients with low prior disease probabilities have specific methodological challenges with patient selection, because the inclusion of a cohort with a proper number of patients with disease suspicion requires a large population to identify sufficient patients with the disease to calculate the sensitivity with adequate precision. Secondly, reference standards for disease detection are almost always performed in hospital settings and often require invasive tests. Therefore, the evaluation of triage tests in non-referred primary care populations can be difficult or unfeasible. It is important to address and analyse these methodological challenges appropriately to develop analytical and/or methodological solutions.

OBJECTIVE OF THIS THESIS

The main objective of this thesis was to study the diagnostic strategies for IBD in children with chronic gastrointestinal symptoms, focusing on the value of testing with faecal calprotectin in primary care.

OUTLINE OF THIS THESIS

In **chapter 2**, we give a systematic overview of the test characteristics for the symptoms, signs, laboratory tests, and combinations of these tests for IBD in children presenting with gastrointestinal symptoms. In **chapter 3**, we present research into the added value of blood markers and faecal calprotectin testing beyond disease-specific symptoms for IBD in referred children, using individual patient data meta-analysis.

The four subsequent chapters then concern the design and data analysis of the DOK study (DOK: *Darm Onderzoek bij Kinderen*; translated as *bowel research in children*), a prospective cohort study with a follow-up period of 12 months. In **chapter 4**, we describe the methodological challenges for diagnostic accuracy studies in primary care and present the design of the DOK study. In **chapter 5**, we determine the diagnostic accuracy of faecal calprotectin in diagnosing IBD in children aged 4 to 18 years and presenting with chronic gastrointestinal symptoms in primary care, by comparing test results with endoscopy or clinical follow-up results. In **chapter 6**, we compare the added diagnostic value of c-reactive protein and faecal calprotectin beyond alarm symptoms to determine the optimal diagnostic test strategy for referral for specialist care in children suspected of IBD. In **chapter 7**, the diagnostic accuracy of a point-of-care test for faecal calprotectin and lactoferrin in primary care is evaluated. Finally, in **chapter 8**, we discuss the implications of our findings for clinical practice, methodology and future studies.

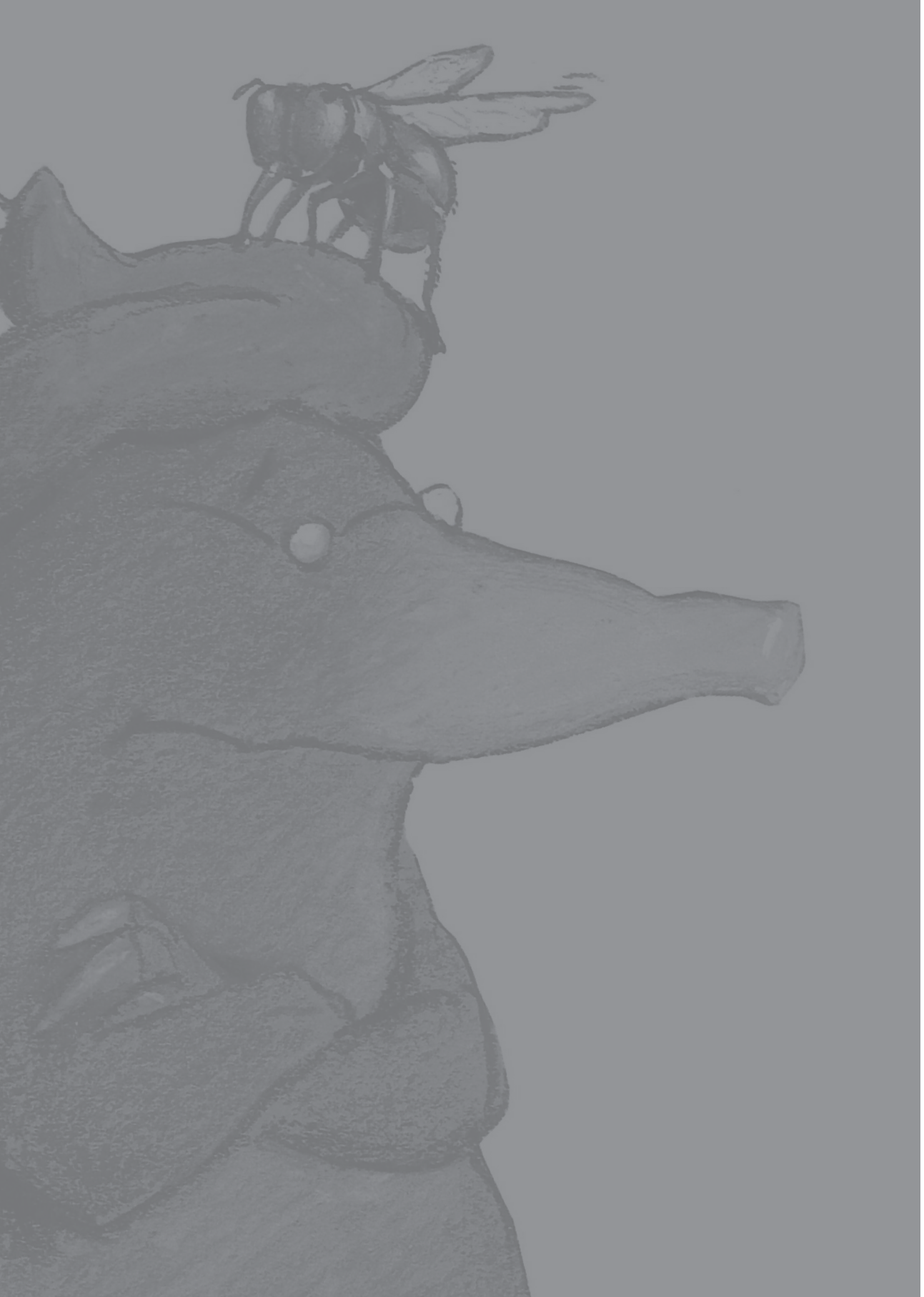
1

REFERENCES

1. Gieteling MJ, Lisman-van Leeuwen Y, van der Wouden JC, Schellevis FG, Berger MY. Childhood nonspecific abdominal pain in family practice: Incidence, associated factors, and management. *Ann Fam Med*. 2011;9(4):337-343.
2. BEACH Program, AIHW General Practice Statistics and Classification Unit. Presentations of abdominal pain in australian general practice. *Aust Fam Physician*. 2004;33(12):968-969.
3. Starfield B, Hoekelman RA, McCormick M, et al. Who provides health care to children and adolescents in the united states? *Pediatrics*. 1984;74(6):991-997.
4. Gieteling MJ, Bierma-Zeinstra SM, Passchier J, Berger MY. Prognosis of chronic or recurrent abdominal pain in children. *J Pediatr Gastroenterol Nutr*. 2008;47(3):316-326.
5. van Tilburg MA, Venepalli N, Ulshen M, Freeman KL, Levy R, Whitehead WE. Parents' worries about recurrent abdominal pain in children. *Gastroenterology Nursing*. 2006;29(1):50-55.
6. Youssef NN, Murphy TG, Langseder AL, Rosh JR. Quality of life for children with functional abdominal pain: A comparison study of patients' and parents' perceptions. *Pediatrics*. 2006;117(1):54-59.
7. Chitkara DK, Rawat DJ, Talley NJ. The epidemiology of childhood recurrent abdominal pain in western countries: A systematic review. *Am J Gastroenterol*. 2005;100(8):1868-1875.
8. Brett T, Rowland M, Drumm B. An approach to functional abdominal pain in children and adolescents. *Br J Gen Pract*. 2012;62(600):386-387.
9. Berger MY, Gieteling MJ, Benninga MA. Chronic abdominal pain in children. *BMJ*. 2007;334(7601):997-1002.
10. Gieteling M, Van Dijk P, De Jonge A, Albeda F, Berger M, Burgers J. *NHG-standaard buikpijn bij kinderen*. *Huisarts Wet*. 2012;55:404-409.
11. van der Zaag-Loonen HJ, Casparie M, Taminiau JA, Escher JC, Pereira RR, Derkx HH. The incidence of pediatric inflammatory bowel disease in the netherlands: 1999-2001. *J Pediatr Gastroenterol Nutr*. 2004;38(3):302-307.
12. Benchimol EI, Fortinsky KJ, Gozdyra P, Van den Heuvel M, Van Limbergen J, Griffiths AM. Epidemiology of pediatric inflammatory bowel disease: A systematic review of international trends. *Inflamm Bowel Dis*. 2011;17(1):423-439.
13. Gupta N, Bostrom AG, Kirschner BS, et al. Presentation and disease course in early-compared to later-onset pediatric crohn's disease. *Am J Gastroenterol*. 2008;103(8):2092-2098.
14. Guariso G, Gasparetto M, Visona Dalla Pozza L, et al. Inflammatory bowel disease developing in paediatric and adult age. *J Pediatr Gastroenterol Nutr*. 2010;51(6):698-707.
15. Kugathasan S, Judd RH, Hoffmann RG, et al. Epidemiologic and clinical characteristics of children with newly diagnosed inflammatory bowel disease in wisconsin: A statewide population-based study. *J Pediatr*. 2003;143(4):525-531.
16. Timmer A, Behrens R, Buderus S, et al. Childhood onset inflammatory bowel disease: Predictors of delayed diagnosis from the CEDATA german-language pediatric inflammatory bowel disease registry. *J Pediatr*. 2011;158(3):467-473. e2.
17. Sawczenko A, Sandhu BK. Presenting features of inflammatory bowel disease in great britain and ireland. *Arch Dis Child*. 2003;88(11):995-1000.
18. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr*. 2014;58(6):795-806.
19. de Ridder L, Rings EH, Escher JC, CBO-werkgroep 'IBD bij kinderen'. Guideline 'diagnosis and treatment of inflammatory bowel disease in children'. *Ned Tijdschr Geneeskd*. 2010;154:A1898.
20. Kim SC, Ferry GD. Inflammatory bowel diseases in pediatric and adolescent patients: Clinical, therapeutic, and psychosocial considerations. *Gastroenterology*. 2004;126(6):1550-1560.
21. Otley AR, Griffiths AM, Hale S, et al. Health-related quality of life in the first year after a diagnosis of pediatric inflammatory bowel disease. *Inflamm Bowel Dis*. 2006;12(8):684-691.
22. Haapamäki J, Roine RP, Sintonen H, Kolho K. Health-related quality of life in paediatric patients with inflammatory bowel disease related to disease activity. *J Paediatr Child Health*. 2011;47(11):832-837.
23. Perminow G, Brackmann S, Lyckander LG, et al. A characterization in childhood inflammatory bowel disease, a new population-based inception cohort from south-eastern norway, 2005-07, showing increased incidence in crohn's disease. *Scand J Gastroenterol*. 2009;44(4):446-456.
24. Van de Vijver E, Schreuder AB, Cnossen WR, Muller Kobold AC, van Rheenen PF, North Netherlands Pediatric IBD Consortium. Safely ruling out inflammatory bowel disease in children and teenagers without referral for endoscopy. *Arch Dis Child*. 2012;97(12):1014-1018.
25. Mack DR, Langton C, Markowitz J, et al. Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics*. 2007;119(6):1113-1119.
26. Cabrera-Abreu JC, Davies P, Matek Z, Murphy MS. Performance of blood tests in diagnosis of inflammatory bowel disease in a specialist clinic. *Arch Dis Child*. 2004;89(1):69-71.
27. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ*. 2013;185(11):E537-44.
28. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: Assessing new tests against existing diagnostic pathways. *BMJ*. 2006;332(7549):1089-1092.
29. Fagerhol MK, Dale I, Anderson T. Release and quantitation of a leucocyte derived protein (L1). *Scand J Haematol*. 1980;24(5):393-398.
30. Røseth A, Fagerhol M, Aadland E, Schjønsby H. Assessment of the neutrophil dominating protein calprotectin in feces: A methodologic study. *Scand J Gastroenterol*. 1992;27(9):793-798.
31. Kristinsson J, Roseth A, Fagerhol MK, et al. Fecal calprotectin concentration in patients with colorectal carcinoma. *Dis Colon Rectum*. 1998;41(3):316-321.
32. Roseth AG, Aadland E, Jahnsen J, Raknerud N. Assessment of disease activity in ulcerative colitis by faecal calprotectin, a novel granulocyte marker protein. *Digestion*. 1997;58(2):176-180.
33. Montalto M, Santoro L, Curigliano V, et al. Faecal calprotectin concentrations in untreated coeliac patients. *Scand J Gastroenterol*. 2007;42(8):957-961.
34. Bunn SK, Bisset WM, Main MJ, Gray ES, Olson S, Golden BE. Fecal calprotectin: Validation as a noninvasive measure of bowel inflammation in childhood inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2001;33(1):14-22.
35. Fagerberg UL, Loof L, Myrdal U, Hansson LO, Finkel Y. Colorectal inflammation is well predicted by fecal calprotectin in children with gastrointestinal symptoms. *J Pediatr Gastroenterol Nutr*. 2005;40(4):450-455.
36. Vaos G, Kostakis ID, Zavras N, Chatzemichael A. The role of calprotectin in pediatric disease. *Biomed Res Int*. 2013;2013:542363.
37. Berni Canani R, Rapacciuolo L, Romano MT, et al. Diagnostic value of faecal calprotectin in paediatric gastroenterology clinical practice. *Dig Liver Dis*. 2004;36(7):467-470.
38. Shulman RJ, Eakin MN, Czyzewski DI, Jarrett M, Ou C. Increased gastrointestinal permeability and

gut inflammation in children with functional abdominal pain and irritable bowel syndrome. *J Pediatr.* 2008;153(5):646-650.

39. Olafsdottir E, Aksnes L, Fluge G, Berstad A. Faecal calprotectin levels in infants with infantile colic, healthy infants, children with inflammatory bowel disease, children with recurrent abdominal pain and healthy children. *Acta Paediatr.* 2002;91(1):45-50.
40. van Rheenen PF, Van de Vijver E, Fidler V. Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: Diagnostic meta-analysis. *BMJ.* 2010;341:c3369.
41. Henderson P, Anderson NH, Wilson DC. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease: A systematic review and meta-analysis. *Am J Gastroenterol.* 2014;109(5):637-645.
42. Degraeuwe PL, Beld MP, Ashorn M, et al. Faecal calprotectin in suspected paediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr.* 2015;60(3):339-346.
43. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: Implications for evidence-based diagnosis. *J Clin Epidemiol.* 2009;62(1):5-12.
44. Schunemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ.* 2008;336(7653):1106-1110.
45. Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM. Quantifying the added value of a diagnostic test or marker. *Clin Chem.* 2012;58(10):1408-1417.
46. Knottnerus JA, Van Weel C, Muris JW. Evaluation of diagnostic procedures. *Br Med J.* 2002;324(7335):477.



CHAPTER 2

NON-INVASIVE TESTS FOR INFLAMMATORY BOWEL DISEASE: A META-ANALYSIS

Gea A Holtman

Yvonne Lisman-van Leeuwen

Johannes B Reitsma

Marjolein Y Berger

Pediatrics. 2016;137(1):1-11.

ABSTRACT

BACKGROUND

The clinical presentation of paediatric inflammatory bowel disease (IBD) is often nonspecific and overlaps with functional gastrointestinal disorders.

OBJECTIVE

To determine the diagnostic accuracy of symptoms, signs, non-invasive tests, and test combinations that can assist the clinician with the diagnosis of IBD in symptomatic children.

METHODS

A literature search via Medline and Embase. Two reviewers independently selected studies reporting on diagnostic accuracy of tests for IBD, with confirmation by endoscopy and histopathology or clinical follow-up, in children with chronic gastrointestinal symptoms. Two reviewers independently extracted data and assessed study quality with the QUADAS-2, an evidence-based quality assessment tool for diagnostic accuracy studies.

RESULTS

Nineteen studies were included (N=2806). Symptoms (abdominal pain, diarrhoea, rectal bleeding, weight loss) had pooled sensitivities ranging from 0.48 to 0.82 and specificities ranging from 0.17 to 0.78. Of all blood markers, C-reactive protein (CRP) (9 studies) and albumin (6 studies) had the best performance with pooled sensitivities of 0.63 (0.51-0.73) and 0.48 (0.31-0.66), respectively, and specificities of 0.88 (0.80-0.93) and 0.94 (0.86-0.98), respectively. Faecal calprotectin (FCal) (10 studies) had a pooled sensitivity of 0.99 (0.92-1.00) and specificity of 0.65 (0.54-0.74). One limitation was that none of the studies was conducted in non-referred children.

CONCLUSIONS

In children whose paediatrician is considering an endoscopy, symptoms are not accurate enough to identify low risk patients in whom endoscopy can be avoided. FCal, CRP and albumin are of potential clinical value, given their ability to select children at low risk (negative FCal test) or high risk (positive CRP or albumin test) for IBD.

INTRODUCTION

Chronic gastrointestinal symptoms in children are a common reason to visit a physician. Differentiation between functional bowel disorders, in which diagnostic testing should be minimized, and inflammatory bowel disease (IBD), which should not be missed, is a diagnostic challenge. The incidence of paediatric IBD is low (5.2 per 100.000 per year),[1] although increasing.[2] Symptoms of IBD are often atypical. Only 25% of children with Crohn's disease present with the classic triad of symptoms: diarrhoea, abdominal pain, and weight loss.[3] Testing all children with chronic gastrointestinal symptoms for IBD is neither necessary nor efficient, particularly not in primary care. Furthermore, to confirm or rule out IBD, an endoscopy is necessary, which is invasive and requires general anaesthesia when performed in children.[4] In contrast, an attitude of "wait and see", may cause unnecessary concerns and loss of wellbeing in children with IBD.[5]

Non-invasive tests for IBD, such as blood markers, faecal markers and ultrasonography, may assist the clinician with this diagnostic dilemma. These tests can be used as a triage-instrument; they assist in safely ruling out existing IBD and in selecting those patients who are candidates for further investigations.[6] Faecal calprotectin (FCal), an inflammatory marker, is extensively studied in several reviews and meta-analysis and has good properties for ruling out IBD in children presenting to the paediatrician with symptoms suggestive of IBD.[7-9] However, a complete overview of the diagnostic accuracy of all symptoms, signs and non-invasive tests is lacking. Moreover, the optimal combination of tests or the additional value of a single test to symptoms or other tests has rarely been studied.

The performance of symptoms, signs, and non-invasive tests may vary between non-referred and referred children due to a different patient mix and underlying disorders, as well as different moments in the course of disease at which patients present or varying reference tests on which a diagnosis is based. The goal of the present study was to systematically review the literature to provide an overview of the accuracy of single symptoms, signs, and non-invasive tests for IBD diagnosed via endoscopy in children presenting with chronic gastrointestinal symptoms in all health care settings. A secondary goal was to present accuracy of test combinations and the added value of a single test to symptoms, signs, or other tests.

METHODS

SEARCH STRATEGY

A literature search for eligible diagnostic studies was conducted in Medline and Embase (from inception to September 18, 2014), using Medical Subject Headings, Emtree terms, and free text words related to child, target condition (IBD), and diagnostic accuracy (Appendix 1). A search strategy was constructed specific for diagnostic accuracy studies based on published search strategies.[10] An information expert assisted the search. In addition, two authors (YLvL, GAH) hand-searched the references of all included full text articles, three systematic reviews,[7-9] and guidelines on paediatric IBD.[4,11] No language restrictions were applied to the searches.

STUDY SELECTION

We identified studies performed in all healthcare settings. Six criteria were used to choose studies: 1) the study population consisted of children with gastrointestinal symptoms suggestive of IBD (studies including healthy control subjects and/or patients with known IBD were excluded); 2) one of the following diagnostic tests was investigated: signs, symptoms, markers (blood, faecal or urinary), or ultrasonography; 3) the reference standard for IBD was endoscopy, including histopathology and/or clinical follow-up; 4) the target condition was IBD; 5) the study design provided information about the association between tests of interests and the presence or absence of IBD; and 6) the study report, or the subsequent data requested, enabled the construction of a 2 x 2 table. Authors were contacted if data for the 2 x 2 table were insufficient or missing.

Two reviewers (YLvL, GAH) independently screened titles and abstracts of all identified articles and assessed full-text articles of each potentially eligible study for inclusion. Disagreement between both the reviewers was resolved by discussion and, if necessary, by a third reviewer (MYB). If the full text of an included study was not available, first or last author was contacted.

DATA EXTRACTION AND QUALITY ASSESSMENT

Two reviewers (YLvL, GAH) independently performed data extraction and quality assessment, using standardized forms. Disagreements between the reviewers were resolved by consensus or by a third reviewer (MYB). The following data were extracted: setting and design; study population; index test; reference standard; prevalence of IBD in the study population; number of patients with Crohn's disease, ulcerative colitis, IBD unclassified, or no IBD, and data for the 2 x 2 table.

Study quality was assessed using an evidence-based quality assessment tool for diagnostic accuracy studies (the QUADAS-2).[12] Scores for low or high risk of bias were allocated to four domains: patient selection, index test, reference standard, and flow and timing (Box 1). In addition, concerns were scored regarding applicability for the first three domains.

DATA SYNTHESIS AND ANALYSIS

Diagnostic 2 x 2 tables were imported in Review Manager 5.0 (RevMan, Cochrane collaboration), and sensitivity, specificity, and corresponding 95% confidence intervals (CIs) were calculated for each symptom, sign, test and test combination. The added value of tests was described when it was reported in the studies. For the meta-analysis, bivariate random effects models were used to calculate pooled estimates of sensitivity, specificity, and likelihood ratios (LRs) when ≥5 studies per index test were included.[13,14] The MIDAS module was used for meta-analysis of diagnostic test accuracy studies in STATA/SE version 12.1 (Stata Corp, College station, TX).

Sources of Heterogeneity

We evaluated whether differences in certain factors could explain identified heterogeneity. These factors included the following: design (cohort or case-control); setting (according to level of selection three settings were defined [children presenting for the first time in primary

Box 1. Signalling questions for scoring the QUADAS 2.

DOMAIN 1: PATIENT SELECTION

Risk of bias

Was a consecutive or random sample of patients enrolled?

Score "yes" if the following words are stated in the article: consecutive, random, or all patients were included between a defined time period.

Was a case-control design avoided?

Score "yes" if a case-control design was avoided.

Did the study avoid inappropriate exclusions?

Score "yes" if studies excluded no other known somatic bowel disorders than IBD or excluded IBD-unclassified.

Applicability concerns

Are there concerns that the included patients and setting do not match the topic of the review question?

Score "low" if the patients were less than 18 years, had symptoms suggestive of IBD and had no previous diagnosis of IBD.

DOMAIN 2: INDEX TEST

Risk of bias

Were the index test results interpreted without knowledge of the results of the reference standard?

Score "yes" if the index test results were interpreted without knowledge of the reference standard.

If a threshold was used, was it pre-specified?

Score "yes" if the threshold was pre-specified.

Applicability concerns

Are there concerns that the index test, its conduct, or interpretation differ from the topic of the review question?

Score "low" if the study provides a clear description of the index test and definition of a positive test result.

DOMAIN 3: REFERENCE STANDARD

Risk of bias

Is the reference standard likely to correctly classify the target condition?

Score "yes" if the reference test is an endoscopy of the upper and lower gastrointestinal tract, including an ileum intubation, histopathology, and/or a clinical follow-up of at least 1 year. (no reference standard is 100% sensitive or specific, but the Porto Criteria state that the ileocolonoscopy including histology is essential)

Were the reference standard results interpreted without knowledge of the results of the index test?
Score "yes" if the reference standard results were interpreted without knowledge of the results of the index test.

Applicability concerns
Are there concerns that the target condition as defined by the reference standard does not match the review question?
Score "low" if the target condition IBD was defined by using an endoscopy or clinical follow-up.

DOMAIN 4: FLOW AND TIMING

Risk of bias
Was there an appropriate interval between index test and reference standard?
Score "yes" if the time period was 1 month or less. It is unlikely that the mucosal inflammation spontaneously disappears in one month.
Did all patients receive a reference standard?
Score "yes" if all patients receive a reference standard.
Did all patients receive the same reference standard?
Score "yes" if all patients receive the same reference standard.
Were all patients included in the analysis?
Score "yes" if all patients included in the study were included in the analysis.

If all answers of signalling questions concerning a domain were "yes", the risk of bias was judged as low. If any signalling question was answered "no" the risk of bias was judged as high. When insufficient data was reported the specific item was classified as "unclear".

care (non-referred, low risk); children referred by their general practitioner (either general practitioner or paediatrician) to a paediatrician or paediatric gastroenterologist for diagnostic evaluation (referred, moderate risk) and children referred by a paediatrician to a paediatric gastroenterologist and endoscopy (referred, high risk)]); number/choice of reference standards (one or two, endoscopy or follow-up); prevalence; and cut-off value of the index test. In case of outliers we evaluated whether bias or specific study characteristics could explain the result. A subgroup analysis (≥5 studies per subgroup) or sensitivity analysis without outliers was performed to evaluate the effect of heterogeneity on test characteristics.

Potential clinical impact
Our goal was to provide more insight in the potential clinical consequences of using the results of the investigated tests. For each test for which we were able to calculate pooled sensitivity and specificity, hypothetical 2 x 2 tables were constructed in 100 children with gastrointestinal symptoms. The number of children with IBD was based on the mean IBD prevalence in the cohort studies included in the meta-analysis. By standardizing the prevalence, it is possible to compare the results of each test. The 2 x 2 tables were based on the pooled estimates of sensitivity and specificity of the index test. Clinical impact was interpreted as follows: children with IBD missed are those with IBD and a negative index test result; the numbers of unnecessary endoscopies are the children without IBD with a positive index test result; and the reduction of patients requiring endoscopy is the total number of patients with a negative index test result. For calculating the latter, we assumed that in the alternative strategy all 100 hypothetical children would undergo endoscopy.

RESULTS

SELECTION, CHARACTERISTICS, AND QUALITY OF STUDIES

The literature search yielded 19 diagnostic studies concerning a total of 2806 children with gastrointestinal symptoms (age range: 3 months to 21 years) of whom 1265 had IBD (Figure 1). The mean prevalence of IBD in the cohort studies was 54% (range: 19% to 82%).[15-28] The characteristics of the 14 cohort studies[15-28] and 5 case-control studies[29-33] are presented in

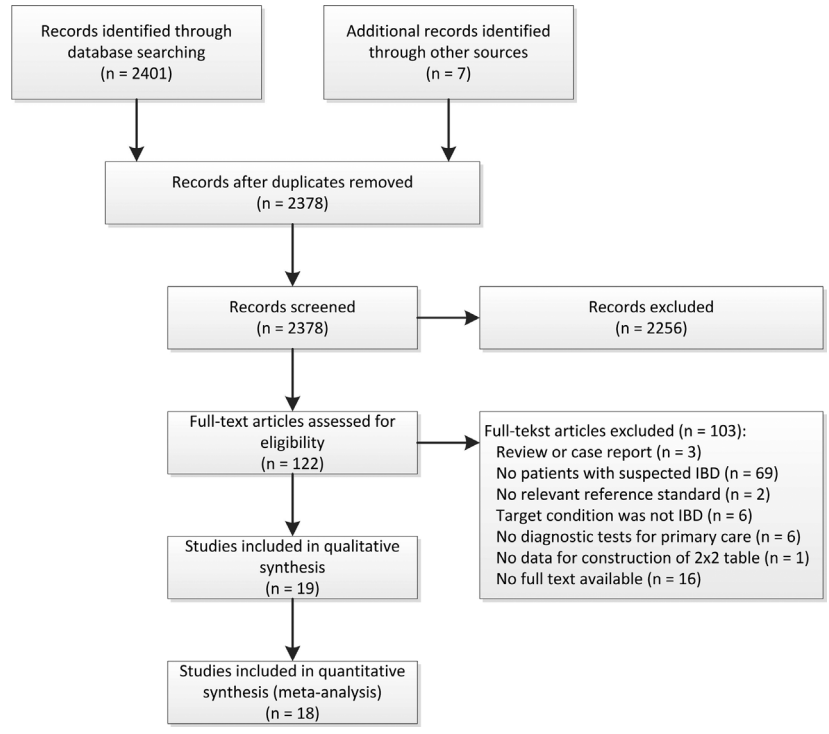


Figure 1. PRISMA flow diagram exhibiting the elimination process for study analysis.
Databases used for literature searching were Medline (n=1619) and Embase (n=782). Four included papers were identified by reference checking.

Table 1. Summary of the methodological assessment of included studies.

	Risk of bias				Applicability concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Cohort Studies							
Ashorn et al, 2009[15]	Unclear	Unclear	Unclear	High	High	Low	Low
Beattie et al, 1995[16]	Low	Low	Unclear	Unclear	Low	Low	Low
Bonnin et al, 2007[17]	Unclear	High	Unclear	Unclear	Low	Low	Low
Cabrera-Abreu et al, 2003[18]	Unclear	High	Unclear	Unclear	Low	Low	Low
Canani et al, 2006[19]	Low	Unclear	High	Unclear	Low	Low	Low
Diamanti et al, 2010[21]	Low	Low	Low	Unclear	Low	Low	Low
Dubinsky et al, 2001[20]	High	Unclear	High	Unclear	High	Low	Low
Fagerberg et al, 2005[22]	High	Unclear	Unclear	High	Low	Low	Low
Khan et al, 2002[23]	High	Low	High	High	Low	Low	Low
Perminow et al, 2009[24]	Unclear	Unclear	High	High	Low	Low	Low
Sabery et al, 2007[25]	High	Low	High	High	High	Low	Low
Sidler et al, 2008[26]	Unclear	High	Low	Unclear	Low	Low	Low
Van de Vijver et al, 2012[27]	Unclear	Low	High	High	Low	Low	Low
Ziech et al, 2014[28]	Low	Unclear	Unclear	High	Low	Low	Low
Case-control studies							
El Chammas et al, 2013[33]	High	Unclear	Unclear	High	Low	Low	Low
Henderson et al, 2012[30]	High	Low	High	High	Low	Low	Low
Leach et al, 2007[32]	High	Unclear	High	Low	Low	Low	Low
Minar et al, 2014[31]	High	Unclear	Unclear	High	Low	Low	Low
Tsampalieros et al, 2011[29]	High	Unclear	Unclear	Unclear	Low	Low	Low

High risk of bias in domain patient selection if the study had no consecutive or random sample of patients, case-control design,[29-33] or inappropriate exclusions.[20,22,23,25,29,32] High risk of bias in domain index test if the index results were interpreted with knowledge of the reference standard, or if the threshold was not pre-specified.[17,18,26] High risk of bias in domain reference standard if the endoscopy did not include an ileum intubation,[19,30,32] the follow-up was less than 12 months,[27] or if the reference standard results were interpreted with knowledge of the index test.[20,23-25,30] High risk of bias in the domain flow and timing if the time period between index test and reference standard was more than 1 month,[15,25,27,30] not all patients receive a reference standard, not all patients receive the same reference standard,[25,27,33] or not all patients were included in the analysis.[15,22-25,28,31] High applicability concerns for the domain patient selection if the study included children aged >18 years.[15,20,25]

Appendix 2. We requested data of incorrect[19] or insufficient[18,20,22,24-28,33] 2 x 2 tables for index tests and additional data were received from five studies.[22,24,26-28] Three cohort studies used two reference standards: endoscopy and follow-up.[25,27,33] None of the studies was performed in non-referred children (low risk). One study included children referred by general practitioners, but the high prevalence of IBD (62%) indicated to us that the children were probably not selected consecutively, and we therefore scored the risk as moderate or high.[24] In four studies, the referred children were at moderate or high risk, [23,25,27,33] in 13 studies, they were at high risk, [7,15-22,26,28,31,32] and in one study the setting was unclear.[29] Two case-control studies reported solely on Crohn's disease.[31,33] Table 1 presents the risk of bias of all studies. Seventeen studies found a high risk of bias in ≥1 domain. On average, the reviewers resolved the disagreement on 2 of 14 items per study (range: 0-5).

DIAGNOSTIC ACCURACY

Diagnostic accuracy measures of all symptoms, signs, tests and test combinations evaluated are presented in Appendix 3A-E. Table 2 presents the results of the meta-analysis of symptoms, signs, and tests evaluated in ≥5 studies. Setting (moderate/high versus high risk), prevalence, and number of reference standards varied little between studies and could not explain heterogeneity in test characteristics of any of the symptoms, signs, or tests evaluated.

Symptoms and signs

The sensitivity and specificity varied substantially between studies for all symptoms (8 studies, Appendix 3A). Study design could not explain heterogeneity. Rectal bleeding had the highest positive likelihood ratio of 2.6 (1.7-4.0).

Blood markers

Sensitivity varied considerably within each blood marker studied (in total, 8 blood markers studied in 13 studies, Appendix 3B). Specificity was fairly homogenous within all blood markers. C-reactive protein (CRP) (cut-off range 3-10 mg/L) was evaluated in nine studies.[16,19,22,24,26,27,29,30,32] Two studies had high sensitivities and high specificities compared to the other studies, in which only specificity was high.[16,29] Heterogeneity could not be explained by differences in study design or cut-off value, nor could we identify specific reasons for bias. Pooled sensitivities for CRP with and without these two outliers were 0.63 (0.51-0.73) and 0.57 (0.46-0.66), respectively, and pooled specificities were 0.88 (0.80-0.93) and 0.84 (0.77-0.89).

Platelet count was evaluated in eight studies.[16,18,19,22,24,26,30,32] A cut-off >400 x 10^9/L yielded lower sensitivities compared to lower cut-off values. The pooled sensitivities with and without studies with a cut-off value <400 x 10^9/L[18,32] were 0.55 (0.36-0.73) and 0.45 (0.28-0.63), respectively, the specificities were 0.88 (0.81-0.93) and 0.91 (0.87-0.94). The study of Beatti et al[16] was identified as an outlier in which sensitivity (0.82) was high with a cut-off of 400 x 10^9/L. Pooled sensitivity and specificity without Beatti et al were 0.37 (0.28-0.47) and 0.92 (0.87-0.95).[19,22,24,26,30]

The nine studies evaluating haemoglobin showed that age/sex-specific cut-off values [22,23,27,30,33] had higher sensitivity than fixed cut-offs.[16,19,20,24] Pooled sensitivity increased from 0.37 (0.24-0.52) to 0.56 (0.46-0.65) when the studies with fixed cut-offs were excluded.

Pooled specificity did not change: 0.90 (0.83-0.94) and 0.87 (0.77-0.93), respectively. For erythrocyte sedimentation rate (ESR) and albumin, the large variation in sensitivity could not be explained by differences in study design or cut-off value. There were no outliers.

Faecal markers

One study[26] reported the diagnostic accuracy of faecal SiooA12, (both sensitivity and specificity: 0.97 [0.83-1.00]). Ten studies investigating FCal (cut-off range: 50-100 µg/g) had high sensitivities (>0.86) with small CIs (Appendix 3C). Only three studies reported false-negative test results.[15,24,30] The specificity in the two case-control studies[30,31] was lower compared to the cohort studies. FCal exhibited a pooled sensitivity and specificity of 0.99 (0.92-1.00) and 0.65 (0.54-0.74) (Table 2), and 1.00 (0.86-1.00) and 0.69 (0.63-0.74), after exclusion of two case-control studies.[30,31]

Urinary markers

One study found that measurement of urinary excretion of cellobiose/mannitol with a cut-off of 0.023 had a sensitivity and specificity of 0.41 (0.22-0.61) and 0.67 (0.41-0.87), respectively (Appendix 3C).[19]

Ultrasonography

The sensitivity of bowel wall thickness of >3 mm and several other parameters measured by using ultrasonography (2 studies) in children with gastrointestinal symptoms ranged from 0.78 to 1.00 and specificity ranged from 0.55 to 0.74 (Appendix 3C).[19,28]

Combinations of tests

Various studies reported on the accuracy of combinations of symptoms and/or tests, but the vast majority of combinations were only studied in a single study (Appendix 3D-E).[16,18-20,23,25-27,30-33] Three combinations were reported in two[23,25,27,29] or three studies.[20,26,27] Five combinations included symptoms, 12 combinations included non-invasive tests, and four combinations included symptoms and non-invasive tests. The two "or" combinations, requiring one of included tests to have positive results, with highest sensitivity (0.97) were: 1) FCal or albumin; 2) haemoglobin or ESR or albumin or platelet count or CRP. The "and" combination, requiring all included tests positive, with highest specificity (0.96) was: haemoglobin and ESR.

Four studies investigated the additional value of a single test in addition to symptoms or other tests. Khan et al[23] found that the addition of haemoglobin and ESR to the finding of rectal bleeding increased sensitivity (from 0.68 to 0.86) and decreased specificity (0.92 to 0.59). Another study found that ESR had no additional value when added to platelet count and hemoglobin.[18] Cellobiose/mannitol in urine had no additional value to the combination of FCal, anti-Saccharomyces cerevisiae antibodies (ASCA), perinuclear anti-neutrophil cytoplasmic antibodies (pANCA), and ultrasonography.[19] Van de Vijver et al[27] investigated the additional value of FCal to the "clinical eye" of the paediatrician. FCal reduced the proportion of IBD-negative endoscopies from 38% to 32% without missing one child with IBD.

Table 2. Pooled estimates of diagnostic performance of symptoms and non-invasive tests for IBD in children.

Variable	Studies	n	Prevalence IBD, % (range)*	Sensitivity (95% CI)	Specificity (95% CI)	LR positive (95% CI)	LR negative (95% CI)	Hypothetical cohort with IBD prevalence of 48%		
								Reduction of endoscopies	IBD missed	Unnecessary endoscopies
Symptoms										
Abdominal pain	6	684	43 (19-62)	0.82 (0.66-0.92)	0.17 (0.12-0.24)	1.0 (0.9-1.1)	1.03 (0.60-1.77)	18	9	43
Diarrhoea	6	684	43 (19-62)	0.76 (0.64-0.85)	0.57 (0.44-0.69)	1.8 (1.4-2.3)	0.42 (0.29-0.59)	42	12	22
Rectal bleeding	7	1280	42 (19-62)	0.57 (0.47-0.66)	0.78 (0.65-0.88)	2.6 (1.7-4.0)	0.55 (0.47-0.65)	62	21	11
Weight loss	6	1173	39 (19-60)	0.48 (0.31-0.66)	0.69 (0.55-0.81)	1.6 (1.1-2.3)	0.74 (0.56-0.99)	61	25	16
Non-invasive tests										
CRP	9	1146	49 (36-62)	0.63 (0.51-0.73)	0.88 (0.80-0.93)	5.1 (2.8-9.4)	0.42 (0.30-0.59)	64	18	6
ESR	11	1434	55 (36-67)	0.66 (0.58-0.73)	0.84 (0.80-0.88)	4.2 (3.3-5.3)	0.41 (0.33-0.50)	60	16	8
Platelet count	8	732	58 (43-62)	0.55 (0.36-0.73)	0.88 (0.81-0.93)	4.7 (2.9-7.7)	0.51 (0.34-0.76)	68	22	6
Haemoglobin	9	1454	50 (36-62)	0.37 (0.24-0.52)	0.90 (0.83-0.94)	3.7 (2.3-5.9)	0.70 (0.57-0.86)	77	30	5
Albumin	6	527	53 (43-62)	0.48 (0.31-0.66)	0.94 (0.86-0.98)	8.3 (3.7-18.7)	0.55 (0.40-0.76)	74	25	3
FCal	10	867	53 (32-82)	0.99 (0.92-1.00)	0.65 (0.54-0.74)	2.8 (2.1-3.7)	0.01 (0.00-0.13)	35	1	18

*Prevalence of IBD based on cohort studies. IBD missed indicates children with IBD and a negative index test result; unnecessary endoscopies indicates children without IBD with a positive index test result; reduction of endoscopies indicates total number of children with a negative index test result. For calculating the latter we assumed that in the alternative strategy all 100 patients would undergo endoscopy. IBD: inflammatory bowel disease, LR, likelihood ratio; CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, FCal: faecal calprotectin.

Clinical impact

Table 2 shows the potential clinical impact of each index test based on their pooled estimates of test accuracy in 100 hypothetical children with gastrointestinal symptoms. Given that prevalence did not explain heterogeneity of sensitivity and specificity in our study, we took the weighted mean prevalence as representative for the population on which our result might be extrapolated. In the hypothetical cohort with an IBD prevalence of 48%, testing with FCal would miss one child with IBD, 18 children without IBD would undergo an unnecessary endoscopy, and the number of children requiring endoscopy compared to a strategy in which all children would undergo endoscopy, would be reduced by 35%. Testing with different blood markers would have missed 16 (ESR) to 30 (haemoglobin) children with IBD; 3 (albumin) to 8 (ESR) children without IBD would have had an endoscopy, the total number of endoscopies would be reduced by 60% (ESR) to 77% (haemoglobin).

DISCUSSION

In this systematic review, we included 19 studies reporting on the diagnostic accuracy of symptoms, signs, non-invasive tests and test combinations for IBD in children with chronic gastrointestinal symptoms. All studies were performed in referred children. The prevalence of IBD was ranging from 19% to 82%. Diagnostic accuracy of individual symptoms and signs was low, with pooled sensitivities for abdominal pain, diarrhoea, rectal bleeding, and weight loss ranging from 0.48 to 0.82 and specificities ranging from 0.17 to 0.78. These findings suggest that in these selected children in whom a paediatrician considers an endoscopy to be indicated, individual signs and symptoms cannot distinguish symptoms caused by IBD from those from other conditions. Therefore, easy tests with few adverse effects (and preferably non-invasive) seem essential in the triage for endoscopy. Use of FCal was best at decreasing the probability of IBD, with a pooled negative likelihood ratio of 0.01 (0.0-0.1). CRP and albumin had pooled positive likelihood ratios sufficiently high to indicate an increase of the probability of IBD that may be of clinical importance (positive likelihood ratios of 5.1 (2.8-9.4) and 8.3 (3.7-18.7), respectively).

Because 17 studies exhibited a high risk of bias in ≥1 domain, the quality of the diagnostic studies was moderate. Three studies used follow-up as the reference standard in children at low risk of IBD instead of performing an endoscopy,[25,27,33] which may lead to missed diagnosis of IBD and overestimation or underestimation of test characteristics. However, no effect on test characteristics was identified by studies that used two reference standards. Case-control design did not influence sensitivity or specificity of symptoms or blood markers, probably because we did not include case-control studies with known IBD case subjects or healthy control subjects, and these factors are acknowledged to overestimate diagnostic accuracy.[34,35]

Blood markers

Heterogeneity in reported sensitivities was high compared to reported specificities. We could explain part of this heterogeneity by the use of different cut-off values. Lower cut-offs for platelet count (<400 x 10^9/L) and age/sex specific cut-offs for haemoglobin increased

sensitivity of both tests. These cut-offs are more appropriate in the triage of IBD, where high sensitivity is important.

CRP exhibited the best overall performance of all the blood markers. This finding remained true after excluding two studies showing extreme high sensitivity and high specificity.[16,29] A narrative review evaluating biomarkers in children and adults also suggests that CRP is the best blood marker to differentiate IBD from functional gastrointestinal disorders.[36] Although CRP exhibited high specificity for IBD, normal CRP levels do not exclude IBD in referred children because the sensitivity is moderate.

Faecal markers

FCal had a high pooled sensitivity (0.99) and a modest pooled specificity (0.65) and is therefore an useful test to rule out IBD in children whose paediatrician suspects IBD and considers endoscopy. Only three out of ten studies reported false-negative results for FCal. These false-negative results might be due to the higher threshold of >100 µg/g instead of >50 µg/g in one study.[15] Moreover, it is possible that in these studies, FCal was measured at an earlier stage of disease compared with other studies.[24,30] Children with IBD at an early stage may not yet have developed a sufficiently elevated calprotectin level. Whether the lower specificity in the case-control studies[30,31] compared with the cohort studies is associated with the selection of patients remains unclear.

The diagnostic accuracy of FCal for paediatric IBD was evaluated in four meta-analysis,[7,8,37,38] one of which was an individual patient data analysis.[38] One of the meta-analyses had methodological limitations and included known IBD case subjects and healthy control subjects, which may lead to overestimation of the diagnostic accuracy.[37] The three high-quality reviews reported the diagnostic accuracy of FCal in children with symptoms suggestive of IBD (average prevalences ranged from 54% to 61%). The pooled sensitivity and specificity reported in the three reviews varied between 0.92 and 0.98, and 0.68 and 0.76, respectively, which are comparable to our results.[7,8,38] The small differences might be due to inclusion of different studies and variations in the 2 x 2 tables. Van Rheenen et al[8] included two studies in which few children were already diagnosed with IBD during faeces sampling,[39,40] Henderson et al[7] excluded studies in which follow-up was used as a reference standard,[27] and Degraeuwe et al[38] included one study in which some of the children had known IBD during faeces sampling.[40] In our review, studies were excluded that included children with known IBD. Moreover, we included one study[31] issued after publication of the latest review.[38]

Faecal S100A12 showed promising results with high sensitivity and specificity. More research is needed to evaluate the diagnostic accuracy of this marker.

Urinary markers

Urinary markers were rarely studied and showed low discriminating power. They provided no added value in combination with other markers.

Ultrasonography

Ultrasonography might be a feasible test: it is non-invasive, easy accessible, and does not involve radiation. The two studies in our review produced different results; sensitivity

ranged from 0.78 to 1.00 and specificity ranged from 0.55 to 0.74. The high specificity of 1.00 is questionable, because only four of the children studied did not have IBD.[28] A previous systematic review regarding imaging in children with IBD recommended that ultrasonography should not be used for the initial diagnosis of paediatric IBD, because of its low accuracy and high inter-operator variability.[41] However, only one of the three included studies evaluated children, and a few children were already diagnosed with IBD when they were included. Two meta-analyses showed that the diagnostic performance of ultrasonography in adults was good; sensitivity ranged from 0.73 to 0.90, and specificity in both reviews was 0.95.[42,43] Before recommending ultrasonography as a triage test for IBD in children, more studies of adequate methodological quality are needed.

Test combinations

Many different test combinations were evaluated, often only in a single study, which hampers comparison of results. Overall, the specificities of combinations of tests were good, whereas sensitivities were less high and heterogeneous. The combined non-invasive test using "or" instead of "and" showed higher sensitivities, which is important for safely excluding IBD. The combined non-invasive test combinations with the highest sensitivity of 0.97 were the combinations "FCal or albumin", and "haemoglobin or ESR or albumin or platelets count or CRP".

Furthermore, the added value of a test to symptoms was rarely studied. One study showed that FCal had added value on the "clinical eye" of the paediatrician.[27] In the latter study, it is unclear how this "clinical eye" incorporated symptoms and blood markers. To investigate the optimal sequential strategy, multivariable logistic regression analyses might be used. A recently published individual patient data analysis constructed a model to predict the probability of having IBD based on FCal and the child's age.[38] The model correctly classified 85.5% of the children, with a sensitivity of 0.81 and specificity of 0.92 (area under the curve: 0.92). Important predictors such as symptoms, signs and other non-invasive tests were not included in this model. Studies or more advanced individual patient data meta-analyses are required to investigate the optimal test strategy for IBD in children with gastrointestinal symptoms. Because of varying IBD prevalence and thresholds for further testing, such strategies might differ between non-referred and referred children.

STRENGTHS AND LIMITATIONS

The strength of the present review is that we evaluated the diagnostic accuracy of all non-invasive tests for IBD in children with gastrointestinal symptoms. Before starting the review, we discussed which non-invasive tests can be reasonably deployed in primary care. The tests should be easy to perform, rapid and applicable in primary care. We therefore excluded tests such as magnetic resonance imaging, computed tomography scans, positron emission tomography, scintigraphy, barium follow-through and serology (e.g. ASCA, pANCA). Although ASCA and pANCA are simple and non-invasive, these tests often produce false-negative results and are therefore not recommended for the triage of IBD.[4] They might be helpful in differentiating between Crohn's disease and ulcerative colitis in children with IBD[4] and should be reserved for specialist settings. A promising faecal marker is faecal lactoferrin; however, this marker was not included in our review because it is only studied in children with known IBD and healthy controls.[44-46] Studies in children with symptoms suggestive for IBD are needed.

Despite our extensive search of Medline and Embase, we identified four publications by hand-searching the references of included publications, reviews and guidelines. This might be due to the search strategy for diagnostic accuracy studies, because these search strategies are not 100% accurate in detecting relevant studies.[48] We chose a pragmatic approach, as the search strategy significantly reduced the number of identified studies. By hand-searching the references, we believe that all relevant studies were included. In addition, we contacted authors about incorrect or insufficient 2 x 2 tables, and this follow-up enabled the construction of optimal 2 x 2 tables of tests.

CLINICAL IMPLICATIONS

In referred children with symptoms suggestive of IBD in whom the paediatrician considers endoscopy, FCal showed to be a sensitive test to safely exclude IBD. Assuming that these children otherwise would have undergone an endoscopy, FCal would reduce the number of endoscopies by 35% at the cost of one missed patient with IBD. By testing for FCal only 18% of the patients without IBD would undergo an invasive procedure because of a false-positive test result. One might consider that missing a child with IBD at this level of care, is unacceptable. Therefore, a sequential strategy of tests might be more adequate. In referred children with a positive FCal result, CRP or albumin may be added because of their low false-positive rate, and consequent reduction of unnecessary endoscopies. However, the predictive value may change when tests are applied sequentially instead of being used in isolation.[49] Future research is needed to investigate sequential strategies.

Children included in studies of this systematic review were all referred children. In 16 of the 19 studies all children underwent an endoscopy. The setting (moderate/high versus high) or the prevalence (range: 19%-82%) did not influence the sensitivity or specificity of the symptoms or tests. Therefore, the results of this systematic review are generalizable to paediatricians or paediatric gastroenterologists who evaluate children in whom they consider endoscopy indicated. The patient population of a paediatricians varies between different healthcare systems.[50] In healthcare systems in the Netherlands, United Kingdom, Scandinavia, Canada, New Zealand, and Australia children can only be seen by a paediatrician or paediatric gastroenterologist if they are referred by a general practitioner. In the United States children can visit their general paediatrician directly. By interpreting the results of our review, one must take generalizability to the intended population into account.

An important result is that none of the studies was performed in non-referred low-risk children. In 2005, a technical report about chronic abdominal pain in children stated that symptoms were not evaluated in non-referred children and blood markers were rarely studied, and only in referred children.[47] Although there are now sufficient number of studies investigating non-invasive tests, it is remarkable that studies in non-referred children are still lacking. Therefore, we could not compare the diagnostic accuracy between non-referred and referred children. Moreover, it is not possible to extrapolate our results to populations of non-referred children. Studies evaluating the accuracy of these tests in non-referred children are urgently needed.

CONCLUSION

This present review provides an overview of symptoms, signs, and non-invasive tests for IBD in children presenting with symptoms suggestive for IBD in whom a paediatrician considers endoscopy indicated. In these children, symptoms alone are insufficient in triage for IBD. FCal, CRP and albumin are of clinical value, given their ability to select children at low risk (negative FCal test) or high risk (positive albumin or CRP test) for IBD. Further research should investigate the accuracy of sequential testing strategies and added values of tests beyond sign and symptoms focusing on FCal, CRP, and albumin. Before tests or a diagnostic strategy can be recommended in non-referred, low-risk children, high-quality studies are needed in this setting.

REFERENCES

1. van der Zaag-Loonen HJ, Casparie M, Taminiau JA, Escher JC, Pereira RR, Derkx HH. The incidence of pediatric inflammatory bowel disease in the netherlands: 1999-2001. *J Pediatr Gastroenterol Nutr*. 2004;38(3):302-307.
2. Benchimol EI, Fortinsky KJ, Gozdyra P, Van den Heuvel M, Van Limbergen J, Griffiths AM. Epidemiology of pediatric inflammatory bowel disease: A systematic review of international trends. *Inflamm Bowel Dis*. 2011;17(1):423-439.
3. Sawczenko A, Sandhu BK. Presenting features of inflammatory bowel disease in great britain and ireland. *Arch Dis Child*. 2003;88(11):995-1000.
4. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr*. 2014;58(6):795-806.
5. Kim SC, Ferry GD. Inflammatory bowel diseases in pediatric and adolescent patients: Clinical, therapeutic, and psychosocial considerations. *Gastroenterology*. 2004;126(6):1550-1560.
6. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: Assessing new tests against existing diagnostic pathways. *BMJ*. 2006;332(7549):1089-1092.
7. Henderson P, Anderson NH, Wilson DC. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease: A systematic review and meta-analysis. *Am J Gastroenterol*. 2014;109(5):637-645.
8. van Rheenen PF, Van de Vijver E, Fidler V. Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: Diagnostic meta-analysis. *BMJ*. 2010;341:c3369.
9. Waugh N, Cummins E, Royle P, et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: Systematic review and economic evaluation. *Health Technol Assess*. 2013;17(55):xv-xix, 1-211.
10. Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *J Clin Epidemiol*. 2006;59(3):234-240.
11. de Ridder L, Rings EH, Escher JC, CBO-werkgroep 'IBD bij kinderen'. Guideline 'diagnosis and treatment of inflammatory bowel disease in children'. *Ned Tijdschr Geneeskd*. 2010;154:A1898.
12. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
13. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.
14. Diaz M. Performance measures of the bivariate random effects model for meta-analyses of diagnostic accuracy. *Comput Stat Data Anal*. 2015;83:82-90.
15. Ashorn S, Honkanen T, Kolho KL, et al. Fecal calprotectin levels and serological responses to microbial antigens among children and adolescents with inflammatory bowel disease. *Inflamm Bowel Dis*. 2009;15(2):199-205.
16. Beattie RM, Walker-Smith JA, Murch SH. Indications for investigation of chronic gastrointestinal symptoms. *Arch Dis Child*. 1995;73(4):354-355.
17. Bonnin Tomas A, Vila Vidal M, Rosell Camps A. Fecal calprotectin as a biomarker to distinguish

between organic and functional gastrointestinal disease. *Rev Esp Enferm Dig*. 2007;99(12):689-693.
18. Cabrera-Abreu JC, Davies P, Matek Z, Murphy MS. Performance of blood tests in diagnosis of inflammatory bowel disease in a specialist clinic. *Arch Dis Child*. 2004;89(1):69-71.
19. Canani RB, de Horatio LT, Terrin G, et al. Combined use of noninvasive tests is useful in the initial diagnostic approach to a child with suspected inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2006;42(1):9-15.
20. Dubinsky MC, Ofman JJ, Urman M, Targan SR, Seidman EG. Clinical utility of serodiagnostic testing in suspected pediatric inflammatory bowel disease. *Am J Gastroenterol*. 2001;96(3):758-765.
21. Diamanti A, Panetta F, Basso MS, et al. Diagnostic work-up of inflammatory bowel disease in children: The role of calprotectin assay. *Inflamm Bowel Dis*. 2010;16(11):1926-1930.
22. Fagerberg UL, Loof L, Myrdal U, Hansson LO, Finkel Y. Colorectal inflammation is well predicted by fecal calprotectin in children with gastrointestinal symptoms. *J Pediatr Gastroenterol Nutr*. 2005;40(4):450-455.
23. Khan K, Schwarzenberg SJ, Sharp H, Greenwood D, Weisdorf-Schindele S. Role of serology and routine laboratory tests in childhood inflammatory bowel disease. *Inflamm Bowel Dis*. 2002;8(5):325-329.
24. Perminow G, Brackmann S, Lyckander LG, et al. A characterization in childhood inflammatory bowel disease, a new population-based inception cohort from south-eastern norway, 2005-07, showing increased incidence in crohn's disease. *Scand J Gastroenterol*. 2009;44(4):446-456.
25. Sabery N, Bass D. Use of serologic markers as a screening tool in inflammatory bowel disease compared with elevated erythrocyte sedimentation rate and anemia. *Pediatrics*. 2007;119(1):e193-9.
26. Sidler MA, Leach ST, Day AS. Fecal S100A12 and fecal calprotectin as noninvasive markers for inflammatory bowel disease in children. *Inflamm Bowel Dis*. 2008;14(3):359-366.
27. Van de Vijver E, Schreuder AB, Cnossen WR, Muller Kobold AC, van Rheenen PF, North Netherlands Pediatric IBD Consortium. Safely ruling out inflammatory bowel disease in children and teenagers without referral for endoscopy. *Arch Dis Child*. 2012;97(12):1014-1018.
28. Ziech ML, Hummel TZ, Smets AM, et al. Accuracy of abdominal ultrasound and MRI for detection of crohn disease and ulcerative colitis in children. *Pediatr Radiol*. 2014;44(11):1370-1378.
29. Tsampalieros A, Griffiths AM, Barrowman N, Mack DR. Use of C-reactive protein in children with newly diagnosed inflammatory bowel disease. *J Pediatr*. 2011;159(2):340-342.
30. Henderson P, Casey A, Lawrence SJ, et al. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease. *Am J Gastroenterol*. 2012;107(6):941-949.
31. Minar P, Haberman Y, Jurickova I, et al. Utility of neutrophil fcgamma receptor I (CD64) index as a biomarker for mucosal inflammation in pediatric crohn's disease. *Inflamm Bowel Dis*. 2014;20(6):1037-1048.
32. Leach ST, Yang Z, Messina I, et al. Serum and mucosal S100 proteins, calprotectin (S100A8/S100A9) and S100A12, are elevated at diagnosis in children with inflammatory bowel disease. *Scand J Gastroenterol*. 2007;42(11):1321-1331.
33. El-Chammas K, Majeskie A, Simpson P, Sood M, Miranda A. Red flags in children with chronic abdominal pain and crohn's disease-a single center experience. *J Pediatr*. 2013;162(4):783-787.
34. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282(11):1061-1066.
35. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174(4):469-476.
36. Vermeire S, Van Assche G, Rutgeerts P. Laboratory markers in IBD: Useful, magic, or unnecessary toys? *Gut*. 2006;55(3):426-431.
37. von Roon AC, Karamountzos L, Purkayastha S, et al. Diagnostic precision of fecal calprotectin for inflammatory bowel disease and colorectal malignancy. *Am J Gastroenterol*. 2007;102(4):803-813.
38. Degraeuwe PL, Beld MP, Ashorn M, et al. Faecal calprotectin in suspected paediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2015;60(3):339-346.
39. Bunn SK, Bisset WM, Main MJ, Gray ES, Olson S, Golden BE. Fecal calprotectin: Validation as a noninvasive measure of bowel inflammation in childhood inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2001;33(1):14-22.
40. Kolho KL, Raivio T, Lindahl H, Savilahti E. Fecal calprotectin remains high during glucocorticoid therapy in children with inflammatory bowel disease. *Scand J Gastroenterol*. 2006;41(6):720-725.
41. Duigenan S, Gee MS. Imaging of pediatric patients with inflammatory bowel disease. *AJR Am J Roentgenol*. 2012;199(4):907-915.
42. Jellema P, van Tulder MW, van der Horst HE, Florie J, Mulder CJ, van der Windt DA. Inflammatory bowel disease: A systematic review on the value of diagnostic testing in primary care. *Colorectal Dis*. 2011;13(3):239-254.
43. Horsthuis K, Bipat S, Bennink RJ, Stoker J. Inflammatory bowel disease diagnosed with US, MR, scintigraphy, and CT: Meta-analysis of prospective studies 1. *Radiology*. 2008;247(1):64-79.
44. Walker TR, Land ML, Kartashov A, et al. Fecal lactoferrin is a sensitive and specific marker of disease activity in children and young adults with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2007;44(4):414-422.
45. Pfefferkorn MD, Boone JH, Nguyen JT, Juliar BE, Davis MA, Parker KK. Utility of fecal lactoferrin in identifying crohn disease activity in children. *J Pediatr Gastroenterol Nutr*. 2010;51(4):425-428.
46. Joishy M, Davies I, Ahmed M, et al. Fecal calprotectin and lactoferrin as noninvasive markers of pediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2009;48(1):48-54.
47. Di Lorenzo C, Colletti RB, Lehmann HP, et al. Chronic abdominal pain in children: A technical report of the american academy of pediatrics and the north american society for pediatric gastroenterology, hepatology and nutrition: AAP subcommittee and NASPGHAN committee on chronic abdominal pain. *J Pediatr Gastroenterol Nutr*. 2005;40(3):249-261.
48. Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *J Clin Epidemiol*. 2011;64(6):602-607.
49. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem*. 2004;50(3):473-476.
50. Thomson S, Osborn R, Squires D, Reed SJ. International profiles of health care systems 2011: Australia, canada, denmark, england, france, germany, iceland, italy, japan, the netherlands, new zealand, norway, sweden, switzerland, and the united states. 2011.

Appendix 1. Full text of search strategy

SEARCH STRATEGY PUBMED

Target condition IBD
("Inflammatory bowel diseases"[MeSH] OR inflammatory bowel disease*[tw] OR IBD[tw] OR Colitis [tw] OR Crohn [tw])

Child
("Child"[Mesh] OR "Adolescent"[Mesh] OR "Infant"[Mesh]OR child*[tw] OR infant*[tw] OR adolescent*[tw] OR pediatric[tw] OR paediatric[tw] OR teenager*[tw])

Diagnostic accuracy
("Sensitivity and Specificity"[Mesh] OR specificit*[tw] OR false negative[tw] OR accura*[tw] OR sensitiv*[tw] OR "reproducibility of results"[MeSH])

SEARCH STRATEGY EMBASE

Target condition IBD
('enteritis'/de OR 'colitis'/exp OR 'inflammatory bowel disease':de,ab,ti OR IBD:de,ab,ti OR colitis:de,ab,ti OR crohn:de,ab,ti)

Child
('child'/exp OR 'newborn'/exp OR 'adolescent'/exp OR child*:de,ab,ti OR infant*:de,ab,ti OR adolescent*:de,ab,ti OR pediatric:de,ab,ti OR paediatric:de,ab,ti OR teenager*:de,ab,ti)

Diagnostic accuracy
('sensitivity and specificity'/exp OR 'predictive value'/exp OR 'receiver operating characteristic'/exp OR sensitiv*:de,ab,ti OR specificit*:de,ab,ti OR accura*:de,ab,ti OR 'false negative':de,ab,ti OR 'reproducibility'/exp) NOT [medline]/lim AND [embase]/lim

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD.

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Ashorn 2009[15]	- Finland - Cohort - referred- high risk	Children and adolescents examined for suspicion of IBD	2.7-19.9	60 (82) 18:36:6	13: celiac disease, allergy, lymphatic adenopathy, extrahepatic portal obstruction, juvenile polyp, haematochezia, haemorrhagic intestinalis NAS, operated intestinal invagination, lumbosacralic meningomyelocele, normal	- FCal *Other tests* - Anti-I2 - OmpW - ASCA	-**FCal** >100 µg/g	Endoscopy and histopathology (upper and lower endoscopy)
Beatti 1995[16]	- United Kingdom - Cohort - referred- high risk	Children who were referred for endoscopic assessment and had two or more of the following symptoms: abdominal pain; diarrhoea; rectal bleeding; weight loss; or mouth ulceration for more than three months. Exclusion: one patient with short duration of symptoms (two weeks) and one because of orofacial granulomatosis.	0.25-18	39 (42.9) 26:13:0	52: polyps, normal (gastro-intestinal symptoms), tuberculosis, indeterminate colitis, lymphoid nodular hyperplasia	- CRP - ESR - platelet count - Hb - albumin *Combined tests* - ≥1 positive: platelet count, Hb, CRP, ESR, or albumin	- **CRP** >5 mg/l - **ESR** >25 mm/h - **Platelet count** >400 x10^9/l - **Hb** <100g/l - **Albumin** <36 g/l	Endoscopy, histopathology and barium follow through (including ileocoloscopy)
Bonnin 2007[17]	- Spain - Retrospective cohort - referred- high risk	Referred children who had faecal calprotectine ordered on presenting with signs and symptoms suggestive of organic pathology (intense abdominal pain, chronic diarrhoea, weight loss, rectal bleeding)	0.4 - 15.3	15 (40.5) 12:3:0	22: 13 functional and 19 organic (non IBD)	- FCal	- **FCal** >50 µg/g	Endoscopy and histopathology

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Cabrera-Abreu 2004[18]	- United Kingdom - Cohort - referred- high risk	Children referred to gastroenterology department presenting with various gastrointestinal and nutritional symptoms compatible with IBD.	1-17.6	103 (67.3) 60:37:6	50: functional, coeliac disease, gastroenteritis, gastro-oesophageal reflux, no identified disorder	- ESR - platelet count *No 2x2 table* - CRP - albumin - Hb *Combined tests* - 1 or 2 abnormal: Hb, platlets, or CRP - ≥1 positive: Hb or platlets	- **ESR** >10 ml/h - **Platelet count** >350 x 10^9 /l - **CRP** >5 mg/l - **Albumin** <35 g/l - **Hb** 1-3 y, <110 g/l; >3 - 6 y, ,117 g/l; >6 -18 y, <120 g/l	Endoscopy, histopathology and radiology

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Canani 2006[19]	- Italy - Cohort - referred- high risk	All children referred for suspected IBD. Exclusion: patients with symptoms or signs mandating a complete work up for IBD (right-lower quadrant mass or perianal disease or haematochezia).	Mean (SD): UC 11.0 (5.0) CD: 14.5 (5.1) Non IBD 11.0 (3.3)	27 (60) 17:10:0	18: functional disorders, food allergy-related intestinal diseases, infectious enter colitis, Mediterranean familial fever	- CRP - ESR - platelet count - Hb - FCal - urinary excretion of cellobiose/ manitol - US *Other tests* - ASCA -pANCA *Combined tests* - All positive: Hb, CRP, and ESR - ≥1 positive: Hb, CRP, or ESR - ≥1 positive: FCal or US	- **CRP** UK - **ESR** >20 mm - **Platelet count** >450 x 10^9 /l - **Hb** <10g/dl - **FCal** >95.3 µg/g - **IP** < 0.023 - **BWUS**: >3.0 mm	Endoscopy, histopathology and radiology for small bowel disease

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Diamanti, 2010[21]	- Italy - Cohort - referred- high risk	Recurrent abdominal pain and altered bowel habit, associated with one or more of the following symptoms: rectal bleeding; failure to thrive; loss of weight; delay of pubertal development; abnormality of physical examination (abdominal mass, perianal disease); positive clinical history (fever of unknown aetiology, arthropathy, erythema nodosum and IBD in a first degree); and altered blood tests (ESR, CRP, full blood count, serum albumin, electrolytes, and liver/renal function).	1-18	117 (59.4) 49:68:0	80: normal colon at histology, a specific colitis, benign lymph nodular hyperplasia	-FCal	- **FCal** >100 µg/g	Endoscopy and histopathology (lower endoscopy including ileocolonoscopy)

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Dubinsky 2001[20]	- Canada - Cohort - referred- high risk	Paediatric patients referred for initial assessment of possible IBD. Exclusion: patients with symptoms (haematochezia) or signs (right lower quadrant mass or perianal disease) mandating a complete workup for IBD.	4-19	54 (42.2) 47:7:0	74: Intestinal inflammation presented (n=21): viral gastroenteritis, acute self-limited bacterial colitis, celiac disease, small bowel bacterial overgrowth, lymphocytic colitis, eosinophilic colitis, peptic ulcer disease Intestinal inflammation absent (n=53): recurrent abdominal pain of childhood, lactose intolerance, rheumatological disorders (uveitis, arthritis, erythema nodocum), choledocolithiasis	- abdominal pain - diarrhoea - family history of IBD - Hb *No 2x2 table* - weight loss - short of stature - ESR - platelet count - albumin - Hb - rheumatological disease *Other tests* - ASCA - pANCA *Combined tests* - ≥1 positive: Hb, platelet count, ESR, or albumin - ≥1 positive: weight loss or short stature	- **ESR** >20 mm/h - **Platelet count** >450 x10^9/l - **Albumin** <35 g/l - **Hb** <10g/dl	Endoscopy, histopathology and small bowel follow through (upper and lower endoscopy)

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Fagerberg 2005[22]	- Sweden - Cohort - referred- high risk	Children (n=40) with gastrointestinal symptoms who were scheduled for colonoscopy to rule out IBD. The decision to perform a colonoscopy was made by a paediatric gastroenterologist and based on the child's medical history, physical examination and laboratory blood tests including full blood count, liver function test, ESR, CRP, orosomucoid and albumin. Exclusion (n=4): two delivered their faecal sample after endoscopy and in two evaluating colonic inflammation was not possible because of incomplete colonoscopy with normal findings.	6.5-17.8	20 (55.6) 10:7:3	16: functional bowel disorders, food intolerance, others Inflamed controls: unspecified proctitis, Juvenile colonic polyposis	- abdominal pain* - diarrhoea* - bloody stools* - ESR* - CRP* - platlet count* - albumin* - orosomucoid* - FCal* *No 2x2 table* - Weight loss* - Hb* - leukocytes* - neutrophil*	- **ESR** ≥16mm/h - **CRP** ≥7mg/L - **Platelet count** >400x x10^9/l - **Albumin** <37g/L - **Orosomucoid** ≥1.15 g/L - **FCal** ≥ 50 µg/g - **Hb** 4-6 y,100-150g/l; 7-10 y, 105-150g/l; 11-15 y, 110-160 g/l; F >15 y, 120-150 g/l; M >15 y, 140-160g/l - **Leukocytes** 4-6 y, 5.5-11.5 10^9/L, 7-10 y, 4.5-10.5 10^9/L; 11-15 y, 4.5-10.0 10^9/L; >15y, 3.5-9.0 10^9/L - **Neutrophils** 4-6 y, 1.8-7.1 10^9/L; 7-10 y, 1.8-6.6 10^9/L; 11-15 y, 2.3-6.8 10^9/L; >15, 1.6-8.0 10^9/L	Endoscopy and histopathology (upper and lower endoscopy including ileocolonoscopy)

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Khan 2002[23]	- USA - Retrospective cohort - referred- moderate/ high risk	Children who had been evaluated for IBD with a complete blood count, ESR and colonoscopy with biopsies during the initial clinical assessment. Exclusion: children without serology. Children with indeterminate colitis (n=15), eosinophilic colitis (n=7) and infectious colitis (n=4) were not included in the statistical analysis	3-18	90 (59.6) 39:51: 15 15 children with indeterminate colitis were not included in analysis	61: symptoms suggestive of IBD	- diarrhoea - pain - rectal bleeding - weight loss - ESR - Hb *Other tests* - pANCA - ASCA *Combined tests* - ≥1 positive: Hb or ESR - ≥1 positive: Hb, ESR, or rectal bleeding	- **ESR** 0-16 y, >15 mm/h; M>16 y, >15 mm/h; F>16 y, >20 mm/h - **Hb** 1-10 y, <10.5 g/dl; >10 y, <11.5 g/dl	Endoscopy, histopathology and follow up

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Perminow 2009[24]	- Norway - Prospective cohort - referred- moderate/ high risk	All general practitioners referring patients to the hospitals for further investigations of suspected symptoms for IBD.	0.8-18	62 (62) 39:19:4	38	- family history of IBD *No 2x2 table* -Abdominal pain* - Diarrhoea* -Rectal bleeding* -pubertal maturity - ESR* - CRP* - Hb* -hematocrit - platelet count* - leukocytes - neutrophil granulocytes - ALAT - alkaline phosphatase - albumin* - FCal*	- **ESR** ≥16mm/h - **CRP** ≥7mg/L - **Hb** <10 g/l - **Platelet count** >400x x10^9/l - **Albumin** <37g/L - **FCal** ≥ 50 µg/g	Endoscopy and histopathology (upper and lower endoscopy according Porto criteria)

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Sabery 2007[25]	- USA - Retrospective cohort - referred- moderate/ high risk	Patients who had IBD serology performed between September 2002 and September 2004 were retrospectively reviewed. Exclusion: patients with incomplete medical charts and unknown or questionable diagnosis	1-21	40 (19) 24:15:1	170: abdominal pain not otherwise specified, gastro esophageal reflux, constipation, IBS, lactose intolerance, nonspecific diarrhoea, food allergies, abdominal migraines, abdominal pain (JRA), bacterial overgrowth, abdominal pain (autism), viral gastroenteritis, anxiety/depression, haemorrhoids, dysmotility, eating disorder, metabolic disorder, milk protein allergy, toddlers diarrhoea, autoimmune enteropathy, adhesions (appendicitis), behavioural incontinence, bezoar, gallstones, hepatitis C, hirschprun, henoch chonlein purpura, hematolic uremic syndrome, intermittent jaundice, meckels, short gut, spondylarthropathy, vasculitis (Churg Strauss)	- abdominal pain - diarrhoea - weight loss - rectal bleeding *No 2x2 table:* - ESR - Hb - arthralgia *Other tests* - ANCA - ASCA - anti-Omp-C *Combined tests* - All positive: Hb and ESR - ≥1 positive: Hb or ESR - All positive: abdominal pain and rectal bleeding - All positive: abdominal pain, diarrhoea and rectal bleeding	- **ESR** >20 mm/h - **Hb** <2 y, <10.5 g/dl; 2-12y, ,<11.5 g/dl; M >12 y, <13.5 g/dl ;F >12 y, <11.7 g/dl	Endoscopy and histopathology or 2-36 months follow up

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Sidler 2008[26]	- Australia - Prospective cohort - referred- high risk	Children and adolescents with gastrointestinal symptoms suggestive for an organic gut disease and required further investigation based on clinical assessment. Gastrointestinal symptoms included chronic diarrhoea over more than 1 month, bloody stools, and abdominal pain occurring for at least a month. Exclusion: children with previously established diagnosis of an organic gastrointestinal disease. Infectious gastroenteritis (including parasites and ova) by at least 2 negative stool cultures. Used nonsteroidal anti-inflammatory agents, antibiotics, or corticosteroids in the preceding 2 weeks.	2.2-16	31 (50.8) 30:1:0	30: functional bowel disorder, reflux esophagitis, eosinophilic gastrointestinal disorders, juvenile polyp, perianal fistula, duodenal ulcer, helicobacter gastritis, iron deficiency anaemia, systemic inflammatory process	- abdominal pain - CRP - ESR - platelet count - albumin - faecal S100A12 - FCal *No 2x2 table* - diarrhoea* - bloody stools* - weight loss* - failure to thrive* *Other tests* - serum S100A12 *Combined tests* - ≥1 positive: chronic diarrhoea, or bloody diarrhoea - ≥1 positive: weight loss, or failure to thrive	-**CRP** ≥3 mg/l -**ESR** ≥15 mm/h -**Platelet count** ≥450 x10^9/l -**Albumin** ≤34 g/l -**Fecal** S100A12 >10 mg/kg -**FCal** ≥50 mg/kg	Endoscopy, histopathology and small bowel follow through (upper and lower endoscopy including ileocoloscopy)

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Van de Vijver 2012[27]	- Nether-Lands - Prospective cohort - referred- moderate/ high risk	Children with persisting diarrhoea (>4 weeks) or recurrent (≥2 episodes in 6 months) abdominal pain and diarrhoea and at least one of the following criteria: rectal blood loss, unintended weight loss or linear growth retardation, peri-anal symptoms (skin tag, fistula, fissure, abscess), anaemia or extra-intestinal manifestations (erythema nodosum, arthritis, uveitis), increased markers of inflammation (ESR; CRP)	6-18	42 (35.9) 24:16:2	75: Infectious diseases: viral gastroenteritis, bacterial gastroenteritis, parasites Nutrition-related diarrhoea: toddler diarrhoea, miscellaneous gastrointestinal disease, constipation, heliocobacter pylori gastritis, meckel's diverticulum, solitary rectal ulcer, spontaneous reduction of invagination, eosinophilic gastroenteritis, haemorrhoids, physical activity induced intestinal ischemia, superior mesenteric artery syndrome, celiac disease Non-organic: gastrointestinal diseases, functional abdominal pain, anorexia nervosa Other: juvenile dermatomyositis, spontaneous recovery, no definite diagnoses	- rectal blood loss - peri-anal symptoms - FCal *No 2x2 table* - weight loss* - linear growth retardation - extra intestinal symptoms* - Hb* - CRP* - ESR* *Combined tests* - All positive: unintended weight loss, and linear growth retardation - All positive: Hb and extra intestinal symptoms -≥1 positive: ESR, or CRP	- **FCal** >50 µg/g - **Hb** <12 y, <7.1 mmol/l; >12y M, <8.1 mmol/l; >12y F, <7.4 mmol/l - **CRP** >10 mg/l - **ESR** >20 mm/h	Endoscopy and histopathology (according Porto criteria) or 6 month follow-up

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Cohort studies								
Study	**Country of origin, Study design, Setting**	**Selection criteria**	**Age range (years)**	**IBD cases (n (%)) CD:UC: IBD-U**	**Non IBD (n): diagnosis**	**Index test**	**cut-off value**	**Reference standard**
Ziech 2014[28]	- Netherlands - Prospective cohort - referred-high risk	Children with suspected IBD. Presenting symptoms were abdominal pain, diarrhoea, haematochezia and weight loss in the previous weeks or months. No child had been diagnosed with or treated for IBD. All were scheduled for upper gastrointestinal tract endoscopy, ileocolonoscopy, abdominal US and MR entero- colonography. Exclusion criteria: age <8 years and ≥18 years and general contraindications for undergoing MR imaging (such as metallic implants and claustrophobia).	10-17	23 (82) 12:10:1	5	-abdominal ultrasound* *Other tests* -MR entero-and colonography -dynamic contrast-enhanced MRI image	- wall thickness (>3mm), layered appearance of bowel wall, and the presence of abdominal lymphadenopathy	Endoscopy and histopathology (lower and upper endoscopy including ileocolonoscopy)

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Case-control studies									
Study	**Country of origin, Setting, Study design**	**Selection criteria cases**	**Selection criteria controls**	**Age range (years)**	**IBD cases (n) CD:UC: IBD-U**	**Controls (n): type**	**Index test**	**cut-off value**	**Reference standard**
El-Chammas 2013[33]	- USA - Case-control - referred- moderate/high risk	Newly diagnosed CD confirmed by bowel mucosal biopsies	Children with abdominal pain for at least 1 month and no evidence of organic disease were considered to have FGIDs (included 41% upper endoscopy; 32% endoscopy; normal biochemical workup and did not change in their diagnosis after a 2-year follow up). Patients with organic disorders other than CD such as eosinophilic, esophagitis, peptic ulcer disease, pancreatitis, cystic fibrosis, lactose intolerance, celiac disease, and ulcerative colitis were excluded.	9-15	128 128:0:0	478: FGIDs	- rectal bleeding - weight loss - anaemia *No 2x2 table* - abdominal pain - blood in stool - joint pain - ESR - albumin - see article for other 24 tests *Combined tests* - All positive: anaemia and rectal bleeding - All positive: anaemia and weight loss - All positive: weight loss and rectal bleeding	CHW laboratory norms	Endoscopy and histopathology (upper and lower endoscopy) or no change in diagnosis during 2 years follow-up

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Case-control studies									
Study	**Country of origin, Setting, Study design**	**Selection criteria cases**	**Selection criteria controls**	**Age range (years)**	**IBD cases (n) CD:UC: IBD-U**	**Controls (n): type**	**Index test**	**cut-off value**	**Reference standard**
Henderson 2012[30]	- Scotland - Case-control - referred- high risk	Incident cases of paediatric IBD by standard clinical, histological, and radiological findings since August 1997 have been collected prospectively. Exclusion criteria: insufficient stool sample provided; aged < 1 year or > 18 years of age on the endoscopy; greater than a 6-months delay between the FCal sample and the endoscopy date; FCal sample taken after endoscopy; any previously known, hospital diagnosed, GI disease; and previous upper or lower GI endoscopy	Patients undergoing both upper and lower endoscopy for the clinical suspicion of bowel inflammation, but where paediatric IBD was excluded (since 2001). Exclusion criteria: insufficient stool sample provided; aged < 1 year or > 18 years of age on the endoscopy; greater than a 6-months delay between the FCal sample and the endoscopy date; FCal sample taken after endoscopy; any previously known, hospital diagnosed, GI disease; and previous upper or lower GI endoscopy	5.2-14.0	91 62:21:8	99: irritable bowel syndrome, non-specific colitis, no pathology identified, post-infectious enteropathy, cow's milk/wheat intolerance, pinworms, allergic enteropathy, celiac disease, miscellaneous	- CRP - ESR - platelet count - Hb - albumin - total WCC - FCal *Combined tests* -≥1 positive: FCal or albumin	- **CRP** >10 mg/L - **ESR** >20 mm/h - **Platelets** 150-450 x10^9/l - **Hb** 1-2 y, 113-141 g/l; 2-6 y, ,115-135 g/l; 6-12 y, 113-155 g/l; M 12-18 y, 130-160 g/l; F 12-18 y, 120-160 - **Albumin** 22-50 g/L - **Total WCC** 1-2 y, 6.0-17.5 x10^9/l; 2-6 y, ,5.0-17.0 x10^9/l; 6-12 y, 4.5-14.5 x10^9/l; 12-18 y, 4.5-13.0 x10^9/l - **FCal** >50 µg/g Possible GI inflammation	Endoscopy and histopathology Control group: Endoscopy, histopathology and minimal 12 months follow up

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Case-control studies									
Study	**Country of origin, Setting, Study design**	**Selection criteria cases**	**Selection criteria controls**	**Age range (years)**	**IBD cases (n) CD:UC: IBD-U**	**Controls (n): type**	**Index test**	**cut-off value**	**Reference standard**
Leach 2007[32]	- Australia - Nested Case-control - referred- high risk	Children with endoscopic and histologic findings consistent with IBD. Inclusion criteria: children with gastrointestinal symptoms suggesting an organic gastrointestinal condition such as IBD or coeliac disease and a clinical indication to proceed endoscopy. Between 2 and 15 years, no current anti-inflammatory medication and no previous diagnosis of gastro intestinal disease.	Children found to have normal endoscopic and histologic findings and with normal systemic markers of inflammation. Children with coeliac disease were excluded. Inclusion criteria: children with gastrointestinal symptoms suggesting an organic gastrointestinal condition such as IBD or coeliac disease and a clinical indication to proceed endoscopy. Between 2 and 15 years, no current anti-inflammatory medication and no previous diagnosis of gastro intestinal disease.	5-15	39 29:4:6	49: 33 non IBD: normal endoscopic and histologic findings and with normal systemic markers of inflammation 16 coeliac diseases were excluded	- CRP - ESR - platelet count - albumin *Other tests* - serum calprotectin - serum S100A12	- **CRP** >1.94 mg/l - **ESR** >7.46 mm/hr - **Platelets** >323 x x10^9/l - **Albumin** <37.9 g/l	Endoscopy and histopathology

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

Case-control studies									
Study	**Country of origin, Setting, Study design**	**Selection criteria cases**	**Selection criteria controls**	**Age range (years)**	**IBD cases (n) CD:UC: IBD-U**	**Controls (n): type**	**Index test**	**cut-off value**	**Reference standard**
Minar 2014[31]	-USA - case-control study - referred- high risk	Newly diagnosed children with CD meeting clinical and histologic criteria for CD. Inclusion criteria: children referred for colonoscopy for the suspicion of IBD.	Children with chronic gastrointestinal symptoms who were referred for colonoscopy for the suspicion of IBD and were found to have normal ileocolonoscopic findings by endoscopy and did not have any chronic evidence of intestinal inflammation on histologic examination. Inclusion criteria: children referred for colonoscopy for the suspicion of IBD.	6-18	26 26:0:0	30	- FCal *Other tests* -ileal FcϒRIA -S100A9 mRNA -PMN CD64	- **FCal** >50 µg/g - **FCal** >250 µg/g	Endoscopy and histopathology (including ileocoloscopy)

Appendix 2. Characteristics of diagnostic studies of symptoms, signs and non-invasive tests for paediatric IBD (continued).

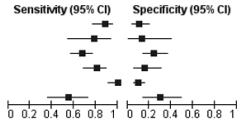
Case-control studies									
Study	**Country of origin, Setting, Study design**	**Selection criteria cases**	**Selection criteria controls**	**Age range (years)**	**IBD cases (n) CD:UC: IBD-U**	**Controls (n): type**	**Index test**	**cut-off value**	**Reference standard**
Tsampalieros 2011[29]	- Canada - Case-control - UK	Newly diagnosed IBD, based on usual clinical, radiologic, endoscopic, and histological criteria. Exclusion: patients with indeterminate colitis.	Those who undergone upper and lower endoscopy with both a normal macroscopic appearance and normal histology of the intestinal mucosa. The most common complaints were abdominal pain, chronic diarrhoea and rectal bleeding.	Mean (SD): UC 12.7 (3.8) CD: 12.5 (3.0) Non IBD 12.3 (3.3)	258 156:102:0	197	- CRP - ESR *No 2x2 table* - Hb - platelets - albumin *Combined tests* -≥1 positive: CRP or ESR	- **CRP** ≥8 mg/l - **ESR** >20 mm/h - **Hb** <6y, <11 g/l; ≥6y <12y, <11.5 g/l; ≥12y F, <12 g/l; ≥12y M, <13 g/l - **Platelets** >450 x10^9/l - **Albumin** <3.4 g/l	Endoscopy and histopathology (upper and lower endoscopy including ileocolonoscopy)

Data represents all patients in each study, instead of those with non-invasive results available. All authors of studies with insufficient or missing data for 2 x 2 tables were contacted, except for Tsampolieros since they had no test results of Hb, platelets, and albumin in the control group. *We received original data to calculate 2 x 2 tables. UK: unknown, IBD: inflammatory bowel disease, CD: Crohn's disease, UC: ulcerative colitis, IBD-U: IBD-unclassified, FGID: functional gastrointestinal disorder, GI: gastrointestinal, CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, Hb: haemoglobin, FCal: faecal calprotectin, US: ultrasound, OmpW: Bacteroides caccae TonB-linked outer membrane protein, ASCA: anti-Saccharomyes cerevisiae antibodies, pANCA: perinuclear anti-neutrophilic antibodies, anti-OmpC: antibodies to outer membrane porin of Escherichia coli, MRI: magnetic resonance imaging, WCC: white cell count, PMN: polymorphonuclear neutrophils

2

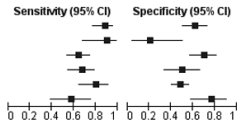
Abdominal pain

Study	TP	FP	FN	TN	Design	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Dubinsky 2001	48	66	6	8	cohort	42.2	0.89 [0.77, 0.96]	0.11 [0.05, 0.20]
Fagerberg 2005	15	13	4	2	cohort	55.6	0.79 [0.54, 0.94]	0.13 [0.02, 0.40]
Khan 2002	61	46	29	15	cohort	59.6	0.68 [0.57, 0.77]	0.25 [0.14, 0.37]
Perminow 2009	50	32	12	6	cohort	62.0	0.81 [0.69, 0.90]	0.16 [0.06, 0.31]
Sabery 2007	40	152	0	18	cohort	19.0	1.00 [0.91, 1.00]	0.11 [0.06, 0.16]
Sidler 2008	17	21	14	9	cohort	50.8	0.55 [0.36, 0.73]	0.30 [0.15, 0.49]



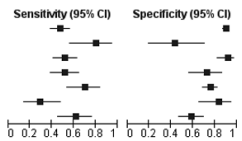
Diarrhea

Study	TP	FP	FN	TN	Design	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Dubinsky 2001	48	28	6	46	cohort	42.2	0.89 [0.77, 0.96]	0.62 [0.50, 0.73]
Fagerberg 2005	18	11	2	3	cohort	55.6	0.90 [0.68, 0.99]	0.21 [0.05, 0.51]
Khan 2002	58	18	32	43	cohort	59.6	0.64 [0.54, 0.74]	0.70 [0.57, 0.81]
Perminow 2009	42	19	20	19	cohort	62.0	0.68 [0.55, 0.79]	0.50 [0.33, 0.67]
Sabery 2007	32	88	8	82	cohort	19.0	0.80 [0.64, 0.91]	0.48 [0.41, 0.56]
Sidler 2008	18	7	13	23	cohort	50.8	0.58 [0.39, 0.75]	0.77 [0.58, 0.90]



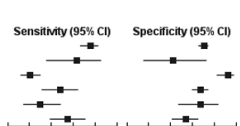
Rectal bleeding

Study	TP	FP	FN	TN	Design	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
El-Chammas 2013	61	48	67	430	case-control		0.48 [0.39, 0.57]	0.90 [0.87, 0.93]
Fagerberg 2005	16	9	4	7	cohort	55.6	0.80 [0.56, 0.94]	0.44 [0.20, 0.70]
Khan 2002	47	5	43	56	cohort	59.6	0.52 [0.41, 0.63]	0.92 [0.82, 0.97]
Perminow 2009	32	10	30	27	cohort	62.0	0.52 [0.39, 0.65]	0.73 [0.56, 0.86]
Sabery 2007	28	41	12	129	cohort	19.0	0.70 [0.53, 0.83]	0.76 [0.69, 0.82]
Sidler 2008	9	5	22	25	cohort	50.8	0.29 [0.14, 0.48]	0.83 [0.65, 0.94]
van de Vijver 2012	26	31	16	44	cohort	35.9	0.62 [0.46, 0.76]	0.59 [0.47, 0.70]



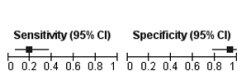
Weight loss

Study	TP	FP	FN	TN	Design	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
El-Chammas 2013	96	143	32	335	case-control		0.75 [0.67, 0.82]	0.70 [0.66, 0.74]
Fagerberg 2005	10	7	6	5	cohort	55.6	0.63 [0.35, 0.85]	0.42 [0.15, 0.72]
Khan 2002	18	5	72	56	cohort	59.6	0.20 [0.12, 0.30]	0.92 [0.82, 0.97]
Sabery 2007	19	56	21	114	cohort	19.0	0.47 [0.32, 0.64]	0.67 [0.59, 0.74]
Sidler 2008	9	10	22	20	cohort	50.8	0.29 [0.14, 0.48]	0.67 [0.47, 0.83]
van de Vijver 2012	23	35	19	40	cohort	35.9	0.55 [0.39, 0.70]	0.53 [0.41, 0.65]



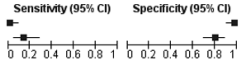
Failure to thrive

Study	TP	FP	FN	TN	Design	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Sidler 2008	6	2	25	28	cohort	50.8	0.19 [0.07, 0.37]	0.93 [0.78, 0.99]



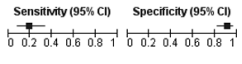
Extra-intestinal symptoms (uveitis, arthritis, arthralgias, erythema nodosum)

Study	TP	FP	FN	TN	Design	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Dubinsky 2001	1	2	53	72	cohort	42.2	0.02 [0.00, 0.10]	0.97 [0.91, 1.00]
van de Vijver 2012	6	15	36	60	cohort	35.9	0.14 [0.05, 0.29]	0.80 [0.69, 0.88]



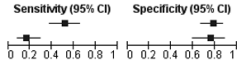
Peri-anal symptoms

Study	TP	FP	FN	TN	Design	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
van de Vijver 2012	8	7	34	68	cohort	35.9	0.19 [0.09, 0.34]	0.91 [0.82, 0.96]



Family history of IBD

Study	TP	FP	FN	TN	Design	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)
Dubinsky 2001	28	16	26	58	cohort	42.2	0.52 [0.38, 0.66]	0.78 [0.67, 0.87]
Perminow 2009	10	9	48	29	cohort	62.0	0.17 [0.09, 0.29]	0.76 [0.60, 0.89]

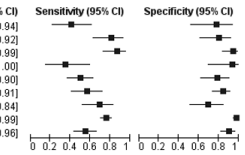


Appendix 3A. Forest plots with sensitivity and specificity of the signs, symptoms, non-invasive tests, and test combinations in individual included studies.

Notes: The studies were ordered by design and cut-off value. 999: cut-off value were age/sex specific. 2x2 tables of Perminow 2009 (abdominal pain, diarrhoea, rectal bleeding, CRP, ESR, platelet count, haemoglobin, albumin, calprotectin), Sidler 2008 (diarrhoea, rectal bleeding, weight loss, failure to thrive), and van de Vijver 2012 (CRP, ESR, haemoglobin, weight loss, extra intestinal symptoms) were based on original data.[24,26,27] The authors of Ziech 2014 provided a 2 x 2 table of ultrasonography.[28] 2 x 2 table of Canani 2006 (FCal) was based on earlier published review,[7] because the sensitivity and specificity were inaccurate in the article.[19] 2 x 2 tables of Fagerberg 2005 (abdominal pain, diarrhoea, rectal bleeding, weight loss, CRP, ESR, platelet count, haemoglobin, albumin, total white cell count, neutrophils, FCal) were based on original data, because we used the outcome IBD instead of organic diseases (we included the two children with organic diseases in the control group).[22] Bonnin 2007 mentioned different numbers in text and table, we used the numbers mentioned in the text.[17] FCal 2 x 2 table of Ashorn 2009 was based on the group without the patients diagnosed before first available faeces sample.[15] Beattie 1995 included three children with indeterminate colitis in the control group.[16]

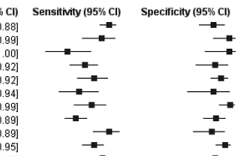
C-reactive protein

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Canani 2006	11	4	16	14	cohort	60.0		0.41 [0.22, 0.61]	0.78 [0.52, 0.94]
Sidler 2008	25	6	6	24	cohort	50.8	3.0	0.81 [0.63, 0.93]	0.80 [0.61, 0.92]
Beatti 1995	34	3	5	49	cohort	42.9	5.0	0.87 [0.73, 0.96]	0.94 [0.84, 0.99]
Fagerberg 2005	7	1	13	15	cohort	55.6	7.0	0.35 [0.15, 0.59]	0.94 [0.70, 1.00]
Perminow 2009	30	8	30	29	cohort	62.0	7.0	0.50 [0.37, 0.63]	0.78 [0.62, 0.90]
van de Vijver 2012	24	12	18	63	cohort	35.9	10.0	0.57 [0.41, 0.72]	0.84 [0.74, 0.91]
Leach 2007	27	10	12	23	case-control		1.94	0.69 [0.52, 0.83]	0.70 [0.51, 0.84]
Tsampalieros 2011	197	5	61	192	case-control		8.0	0.76 [0.71, 0.81]	0.97 [0.94, 0.99]
Henderson 2012	48	8	39	77	case-control		10.0	0.55 [0.44, 0.66]	0.91 [0.82, 0.96]



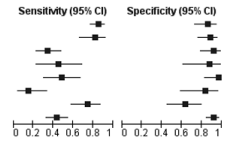
Erythrocyte Sedimentation Rate

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Cabrera-Abreu 2004	84	11	19	39	cohort	67.3	10.0	0.82 [0.73, 0.89]	0.78 [0.64, 0.88]
Sidler 2008	23	2	8	28	cohort	50.8	15.0	0.74 [0.55, 0.88]	0.93 [0.78, 0.99]
Fagerberg 2005	8	1	12	15	cohort	55.6	16.0	0.40 [0.19, 0.64]	0.94 [0.70, 1.00]
Perminow 2009	31	6	23	23	cohort	62.0	16.0	0.57 [0.43, 0.71]	0.79 [0.60, 0.92]
van de Vijver 2012	28	11	14	64	cohort	35.9	20.0	0.67 [0.50, 0.80]	0.85 [0.75, 0.92]
Canani 2006	14	4	13	14	cohort	60.0	20.0	0.52 [0.32, 0.71]	0.78 [0.52, 0.94]
Beatti 1995	25	3	14	49	cohort	42.9	25.0	0.64 [0.47, 0.79]	0.94 [0.84, 0.99]
Khan 2002	44	12	46	49	cohort	59.6	999.0	0.49 [0.36, 0.60]	0.80 [0.68, 0.89]
Leach 2007	32	8	7	25	case-control		7.46	0.82 [0.66, 0.92]	0.76 [0.58, 0.89]
Henderson 2012	58	9	29	74	case-control		20.0	0.67 [0.56, 0.76]	0.89 [0.80, 0.95]
Tsampalieros 2011	193	41	65	156	case-control		20.0	0.75 [0.69, 0.80]	0.79 [0.73, 0.85]



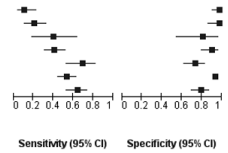
Platelet count

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Cabrera-Abreu 2004	88	7	15	43	cohort	67.3	350.0	0.85 [0.77, 0.92]	0.86 [0.73, 0.94]
Beatti 1995	32	6	7	46	cohort	42.9	400.0	0.82 [0.66, 0.92]	0.88 [0.77, 0.96]
Perminow 2009	21	3	40	35	cohort	62.0	400.0	0.34 [0.23, 0.48]	0.92 [0.79, 0.98]
Fagerberg 2005	9	2	11	14	cohort	55.6	400.0	0.45 [0.23, 0.68]	0.88 [0.62, 0.98]
Sidler 2008	15	1	16	29	cohort	50.8	450.0	0.48 [0.30, 0.67]	0.97 [0.83, 1.00]
Canani 2006	4	3	23	15	cohort	60.0	450.0	0.15 [0.04, 0.34]	0.83 [0.59, 0.96]
Leach 2007	29	12	10	21	case-control		323.0	0.74 [0.58, 0.87]	0.64 [0.45, 0.80]
Henderson 2012	37	7	48	83	case-control		450.0	0.44 [0.33, 0.55]	0.92 [0.85, 0.97]



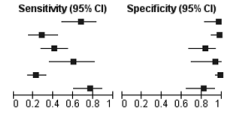
Haemoglobin

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Beatti 1995	8	5	31	47	cohort	42.9	10.0	0.21 [0.09, 0.36]	0.90 [0.79, 0.97]
Canani 2006	6	5	21	13	cohort	60.0	10.0	0.22 [0.09, 0.42]	0.72 [0.47, 0.90]
Dubinsky 2001	6	2	48	72	cohort	42.2	10.0	0.11 [0.04, 0.23]	0.97 [0.91, 1.00]
Perminow 2009	13	1	49	37	cohort	62.0	10.0	0.21 [0.12, 0.33]	0.97 [0.86, 1.00]
Fagerberg 2005	8	3	12	13	cohort	55.6	999.0	0.40 [0.19, 0.64]	0.81 [0.54, 0.96]
Khan 2002	37	6	53	55	cohort	59.6	999.0	0.41 [0.31, 0.52]	0.90 [0.80, 0.96]
van de Vijver 2012	29	20	13	55	cohort	35.9	999.0	0.69 [0.53, 0.82]	0.73 [0.62, 0.83]
El-Chammas 2013	69	29	59	449	case-control			0.54 [0.45, 0.63]	0.94 [0.91, 0.96]
Henderson 2012	57	19	32	72	case-control		999.0	0.64 [0.53, 0.74]	0.79 [0.69, 0.87]



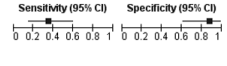
Albumin

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Sidler 2008	21	1	10	29	cohort	50.8	34.0	0.68 [0.49, 0.83]	0.97 [0.83, 1.00]
Beatti 1995	11	1	28	51	cohort	42.9	36.0	0.28 [0.15, 0.45]	0.98 [0.90, 1.00]
Perminow 2009	24	6	35	30	cohort	62.0	37.0	0.41 [0.28, 0.54]	0.83 [0.67, 0.94]
Fagerberg 2005	12	1	8	15	cohort	55.6	37.0	0.60 [0.36, 0.81]	0.94 [0.70, 1.00]
Henderson 2012	20	1	69	82	case-control		33.0	0.22 [0.14, 0.33]	0.99 [0.93, 1.00]
Leach 2007	30	6	9	27	case-control		37.9	0.77 [0.61, 0.89]	0.82 [0.65, 0.93]



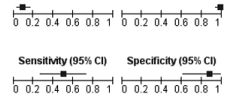
Orosomucoid

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Fagerberg 2005	7	2	13	14	cohort	55.6	1.15	0.35 [0.15, 0.59]	0.88 [0.62, 0.98]



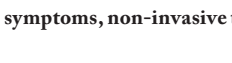
Total White Cell Count

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Fagerberg 2005	8	2	12	14	cohort	55.6	999.0	0.40 [0.19, 0.64]	0.88 [0.62, 0.98]
Henderson 2012	8	1	81	90	case-control		999.0	0.09 [0.04, 0.17]	0.99 [0.94, 1.00]



Neutrophils

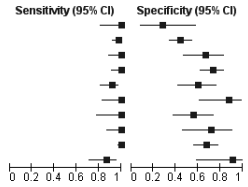
Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Fagerberg 2005	10	2	10	14	cohort	55.6	999.0	0.50 [0.27, 0.73]	0.88 [0.62, 0.98]



Appendix 3B. Forest plots with sensitivity and specificity of the signs, symptoms, non-invasive tests, and test combinations in individual included studies.

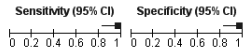
Fecal Calprotectin

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Minar 2014	18	10	0	4	case-control		50.0	1.00 [0.81, 1.00]	0.29 [0.08, 0.58]
Henderson 2012	89	55	2	44	case-control		50.0	0.98 [0.92, 1.00]	0.44 [0.34, 0.55]
Sidler 2008	31	10	0	20	cohort	50.8	50.0	1.00 [0.89, 1.00]	0.67 [0.47, 0.83]
van de Vijver 2012	42	20	0	55	cohort	35.9	50.0	1.00 [0.92, 1.00]	0.73 [0.62, 0.83]
Perminow 2009	54	14	5	21	cohort	62.0	50.0	0.92 [0.81, 0.97]	0.60 [0.42, 0.76]
Fagerberg 2005	20	2	0	14	cohort	55.6	50.0	1.00 [0.83, 1.00]	0.88 [0.62, 0.98]
Bonnin 2007	15	14	0	18	cohort	32.0	50.0	1.00 [0.78, 1.00]	0.56 [0.38, 0.74]
Canani 2006	27	5	0	13	cohort	60.0	95.3	1.00 [0.87, 1.00]	0.72 [0.47, 0.90]
Diamanti 2010	117	26	0	54	cohort	60.0	100.0	1.00 [0.97, 1.00]	0.68 [0.56, 0.78]
Ashorn 2009	32	1	5	10	cohort	82.0	100.0	0.86 [0.71, 0.95]	0.91 [0.59, 1.00]



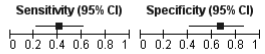
Fecal S100A12

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Sidler 2008	30	1	1	30	cohort	50.8	10.0	0.97 [0.83, 1.00]	0.97 [0.83, 1.00]



Small Intestinal Permeability

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Canani 2006	11	6	16	12	cohort	60.0	0.023	0.41 [0.22, 0.61]	0.67 [0.41, 0.87]



Ultrasonography

Study	TP	FP	FN	TN	Design	Prevalence	Cut off	Sensitivity (95% CI)	Specificity (95% CI)
Canani 2006	20	4	7	14	cohort	60.0	3.0	0.74 [0.54, 0.89]	0.78 [0.52, 0.94]
Ziech 2014	11	0	9	4	cohort		3.0	0.55 [0.32, 0.77]	1.00 [0.40, 1.00]

Appendix 3C. Forest plots with sensitivity and specificity of the signs, symptoms, non-invasive tests, and test combinations in individual included studies.

Abdominal pain and rectal bleeding

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Sabery 2007	29	36	11	134	0.72 [0.56, 0.85]	0.79 [0.72, 0.85]

Abdominal pain and diarrhea and rectal bleeding

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Sabery 2007	26	20	14	150	0.65 [0.48, 0.79]	0.88 [0.82, 0.93]

Rectal bleeding and weight loss

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
El-Chammas 2013	115	182	13	296	0.90 [0.83, 0.94]	0.62 [0.57, 0.66]

Weight loss or failure to thrive

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Dubinsky 2001	12	5	42	69	0.22 [0.12, 0.36]	0.93 [0.85, 0.98]
Sidler 2008	15	12	16	18	0.48 [0.30, 0.67]	0.60 [0.41, 0.77]
van de Vijver 2012	23	35	19	40	0.55 [0.39, 0.70]	0.53 [0.41, 0.65]

Chronic diarrhea or rectal bleeding

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Sidler 2008	20	7	11	23	0.65 [0.45, 0.81]	0.77 [0.58, 0.90]

Haemoglobin and weight loss

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
El-Chammas 2013	106	158	22	320	0.83 [0.75, 0.89]	0.67 [0.63, 0.71]

Haemoglobin and rectal bleeding

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
El-Chammas 2013	100	72	28	406	0.78 [0.70, 0.85]	0.85 [0.81, 0.88]

Haemoglobin or Erythrocyte Sedimentation Rate or rectal bleeding

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Khan 2002	77	20	13	41	0.86 [0.77, 0.92]	0.67 [0.54, 0.79]

Haemoglobin or extra intestinal symptoms

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
van de Vijver 2012	35	35	7	40	0.83 [0.69, 0.93]	0.53 [0.41, 0.65]

Haemoglobin and C-reactive protein and Erythrocyte Sedimentation Rate

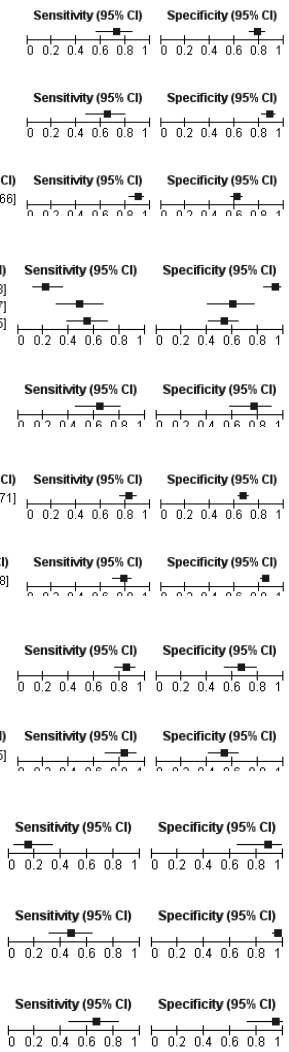
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Canani 2006	4	2	23	16	0.15 [0.04, 0.34]	0.89 [0.65, 0.99]

Haemoglobin and Erythrocyte Sedimentation Rate

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Sabery 2007	19	7	21	163	0.47 [0.32, 0.64]	0.96 [0.92, 0.98]

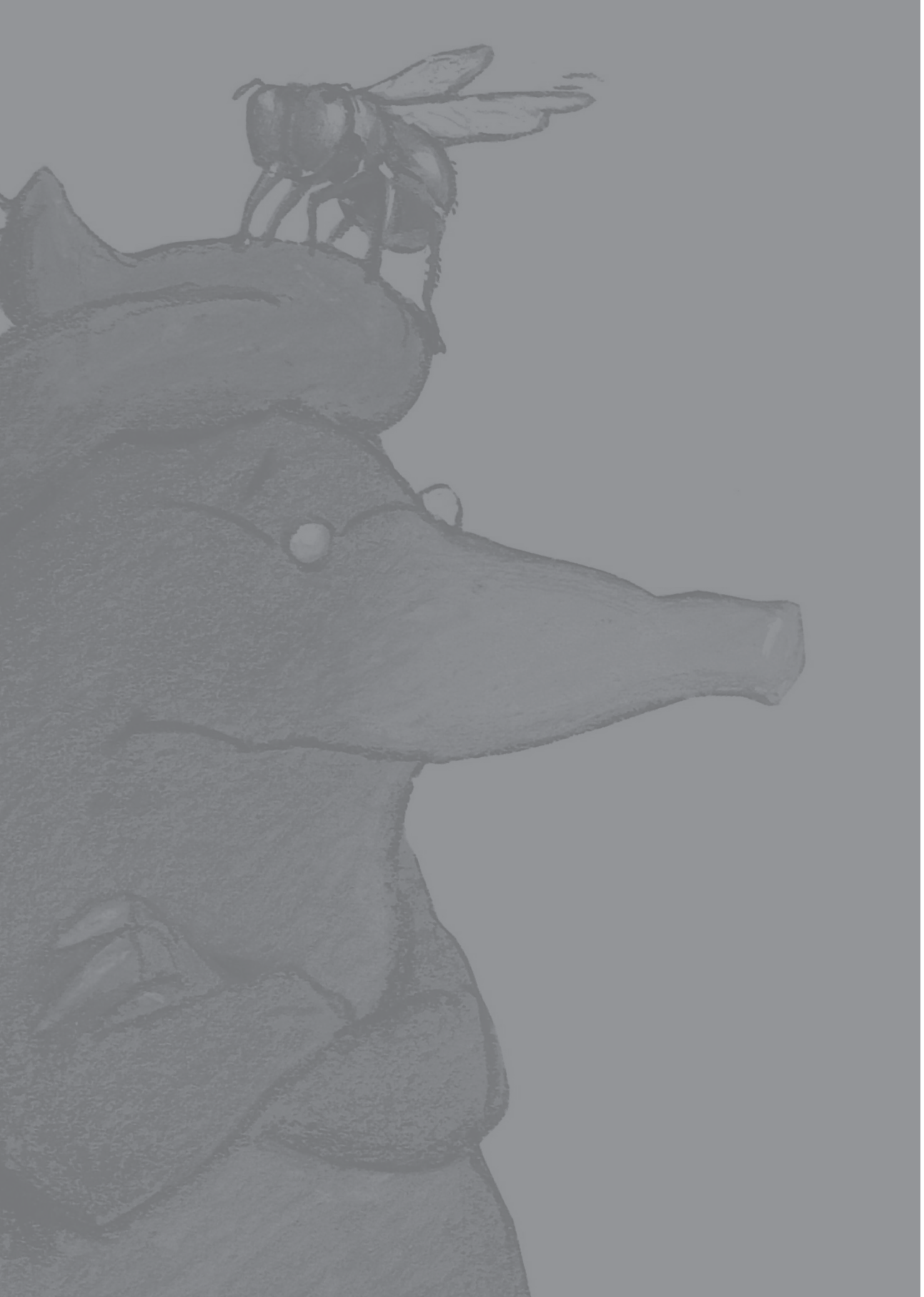
Fecal Calprotectin and Bowel Wall Thickness Ultrasound Measurement

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Canani 2006	18	1	9	17	0.67 [0.46, 0.83]	0.94 [0.73, 1.00]



Appendix 3D. Forest plots with sensitivity and specificity of the signs, symptoms, non-invasive tests, and test combinations in individual included studies.

Appendix 3E. Forest plots with sensitivity and specificity of the signs, symptoms, non-invasive tests, and test combinations in individual included studies.



CHAPTER 3

ADDED VALUE OF LABORATORY MARKERS TO SYMPTOMS FOR INFLAMMATORY BOWEL DISEASE, AN INDIVIDUAL PATIENT DATA META-ANALYSIS OF 1120 PAEDIATRIC PATIENTS

Gea A Holtman, Yvonne Lisman-van Leeuwen,
Andrew S Day, Ulrika L Fagerberg,
Paul Henderson, Stevan Leach,
Gøri Perminow, David Mack,
Patrick F van Rheenen, Els van de Vijver,
David C Wilson, Johannes B Reitsma,
Marjolein Y Berger.

Submitted

ABSTRACT

OBJECTIVE
To evaluate the added diagnostic value of commonly used laboratory markers in addition to symptoms when diagnosing paediatric inflammatory bowel disease using individual patient data from published studies.

DESIGN
An individual patient data meta-analysis. Laboratory markers were added as a single test to a basic prediction model based on symptoms. Outcome measures were improvement of discrimination by adding markers as a single test and improvement of risk classification of paediatric patients by adding the best marker.

DATA SOURCES
A literature search of Medline and Embase.

ELIGIBILITY CRITERIA FOR SELECTING STUDIES
Studies evaluating the diagnostic accuracy of more than one blood marker and faecal calprotectin for paediatric inflammatory bowel disease, confirmed by endoscopy and histopathology or clinical follow-up, in paediatric patients with chronic gastrointestinal symptoms.

RESULTS
Of the 15 eligible studies, authors of 8 studies (n=1120) provided their datasets. All blood markers and faecal calprotectin individually significantly improved the discrimination between paediatric patients with and without IBD, when added to symptoms. The best marker, faecal calprotectin, improved the area under the curve of symptoms by 0.26 (95% CI 0.21-0.31). The second best marker, erythrocyte sedimentation rate, improved the area under the curve of symptoms by 0.16 (95% CI 0.11-0.21). When faecal calprotectin was added to the model, the proportion of paediatric patients without IBD correctly classified as low risk of IBD increased from 33% to 91%. The proportion of paediatric patients with IBD assigned to the low risk of IBD decreased from 16% to 9%. The proportion of the total number of patients assigned to the intermediate risk decreased from 55% to 6%.

CONCLUSIONS
In a hospital setting, faecal calprotectin added the most diagnostic value to symptoms compared to commonly used blood markers. Adding faecal calprotectin to the diagnostic work-up of paediatric patients with symptoms suggestive of inflammatory bowel disease considerably decreased the number of patients in the group in which challenges in clinical decision making are most prevalent.

INTRODUCTION

It is a diagnostic challenge to differentiate between paediatric patients with inflammatory bowel disease (IBD) and functional gastrointestinal disorders, such as irritable bowel syndrome. Unnecessary invasive diagnostic testing and endoscopy need to be balanced against the risk of missing or delaying a diagnosis of IBD. The diagnostic work-up of children and adolescents with gastrointestinal symptoms starts with history and physical examination. Endoscopy is needed to make a definitive diagnosis of IBD, but this is an invasive and unpleasant procedure, especially in paediatric patients.[1] The key question therefore is whether commonly used blood markers or faecal calprotectin improves the accuracy of the diagnostic work-up beyond the finding of history and physical examination to select children for endoscopy.[2] Information on the added values of tests would help the clinician in choosing those tests that are most appropriate and how to correctly interpret the results.

Recently, we published a meta-analysis that provided an overview of the accuracy of signs, symptoms, tests and test combinations for diagnosing IBD in paediatric patients presenting with symptoms suggestive for IBD in whom a paediatrician could consider endoscopy.[3] This meta-analysis was based on published data only and it was therefore not possible to evaluate the added value of tests beyond signs and symptoms. Moreover, the various combinations of tests results were often evaluated only in a single study, and therefore limited information was available on how robust these results were.

High quality evidence for the added value of tests on symptoms can be achieved by using the individual patient data (IPD) from all relevant studies. In this IPD meta-analysis, we evaluate the added diagnostic value of commonly used blood markers and faecal calprotectin on top of signs and symptoms for diagnosing IBD in symptomatic children and adolescents.

METHODS

SEARCH STRATEGY
We searched Medline and Embase from inception till September 18 2014 to identify diagnostic studies which evaluated >1 laboratory test for IBD in paediatric patients with symptoms suggestive for IBD. We used the same literature search as in our recently published meta-analysis[3] that incorporated indexing terms and free text words related to child, target condition IBD and diagnostic accuracy (Appendix 1). Additionally, we hand searched references of full text articles, published reviews and guidelines on paediatric IBD.[1,4-8] No language restrictions were applied.

SELECTION CRITERIA
Two independent reviewers (YvL, GAH) identified and selected eligible studies. All studies examining the diagnostic accuracy of >1 laboratory tests (blood markers or faecal calprotectin) for a diagnosis of IBD were eligible for inclusion. IBD had to be confirmed or rejected by histopathological analysis of biopsies retrieved at endoscopic examination or rejected by the absence of symptoms at clinical follow-up. We included studies that evaluated children

or adolescents, aged 0 to 18 years, with gastrointestinal symptoms suggestive of IBD. We excluded studies that included healthy controls and/or patients with known IBD.

IPD DATASET, DATA EXTRACTION, AND QUALITY ASSESSMENT

We contacted the corresponding author of eligible studies and invited them to share their datasets. In case of non-response, we sent two reminder e-mails. If we had no response after the third e-mail, then the study was excluded from our analysis. From the published reports, two reviewers (YvL, GAH) independently abstracted information on country, study design, setting, and age. In addition the following IPD from each included study was requested: final diagnosis (IBD/no IBD), levels of laboratory tests (blood markers [C-reactive protein, erythrocyte sedimentation rate, platelet count, albumin, haemoglobin] and faecal calprotectin), and, if available, information on the presence of symptoms (abdominal pain, diarrhoea, rectal bleeding, and weight loss). These IPD were compared with the published results. Discrepancies were discussed with the authors and corrected.

Two reviewers (YvL, GAH) independently assessed the risk of bias and concerns for applicability using the QUADAS-2 instrument.[9] The risk of bias and quality of our own study was assessed by two independent reviewers (DCW, PH).[10] This instrument consists of four domains: patient selection, index test, reference standard and flow and timing. Disagreements between reviewers were resolved by consensus or, if necessary, by a third reviewer (MYB).

STATISTICAL ANALYSIS

We used a two-step approach in this IPD meta-analysis to evaluate the discriminative ability of single laboratory markers and the added value to symptoms. In the first step, the results were calculated in each of the individual studies. In the second step, the results were meta-analysed.

Discrimination of markers

In the first step, we determined the discriminative ability of single laboratory markers by calculating the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals (CI) for each dataset. In the second step, we calculated the pooled AUC with 95% CI using the random effects generic inverse variance model.[11]

Added value of markers

First, we developed a common basic model of symptoms considered predictive for IBD (dichotomous dependent variable) using logistic regression analysis in each dataset. The symptoms were abdominal pain, diarrhoea, and rectal bleeding. Other alarming symptoms (e.g. involuntary weight loss, peri-anal lesions, growth failure) were not included in the basic model, because these were not available for all studies. To estimate the added predictive value of single laboratory markers we added these variables as continuous variables to the basic symptoms model. The difference in AUC (delta AUC) with 95% CI between the basic model and the different extended models with a single laboratory marker, was calculated for each dataset using the method of Delong.[12,13] In the second step, a pooled estimate and 95% CI of the delta AUC was calculated by the generic inverse variance method.[11] Moreover, a forest plot was constructed to visualize the delta AUC of each dataset and the heterogeneity between datasets. When the results were heterogeneous, random-effects models were used.

Improvement in diagnostic risk classification

In order to provide more insight in how the paediatric patients were classified by using the basic model and the shift in classification after adding the overall best marker, we constructed a reclassification table. The predicted probability of IBD in all paediatric patients was calculated in each dataset for both models. We defined two threshold probabilities, one below which a paediatrician decides not to perform endoscopy (probability <35%), and one above which a paediatrician decides to perform endoscopy (probability >60%). Therefore three risk groups were created: low risk (predicted probabilities <35%), intermediate risk (predicted probabilities 35-60%) and high risk of IBD (predicted probabilities >60%). The two thresholds probabilities were used to calculate 2 x 2 tables for the basic model and basic model with the best marker in each dataset. The sensitivities and specificities in each dataset were pooled with bivariate random effects models.[14] These pooled sensitivities and specificities and the median prevalence of IBD were used to construct a reclassification table of 100 hypothetical paediatric patients with three relevant risk groups of IBD.

Missing data

If a specific marker was not evaluated in a single study, this dataset was not included when calculating a pooled estimate of that marker. If one or more of the three key symptoms was not evaluated in a study, this study was not included in the evaluation of the added value of the various markers. In case of sporadic missing data, we used multiple imputations (fully condition specification, predictive mean matching, 20 iterations, 5 datasets), with the following variables as predictors: all symptoms (if present), all laboratory markers, and diagnosis.[15,16] We used Rubin's rule to calculate the pooled AUC.[17]

Statistical analyses were performed with IBM SPSS version 20.0.0 (IBM corp., Armonk, New York, USA), STATA/SE 13 (Stata Corp, College station, TX, USA), and SAS 9.2 (SAS institute, Cary, NC, USA).

RESULTS

SELECTION OF STUDIES

Of the 2378 studies identified from the literature search, 15 diagnostic studies were eligible (Figure 1). From seven studies (n=1384 patients) the IPD were not obtained, because three authors did not respond to e-mails,[18-20] the data was no longer available,[21-23] or the author declined to share data.[24] The mean prevalence of IBD in the six excluded cohort studies was 49% (range 19%-67%).[18-23] One excluded study used a case-control design in symptomatic paediatric patients.[24] Four of the seven excluded studies reported on symptoms and blood markers,[20,21,23,24] two reported on blood markers only,[18,22] and one study on blood markers and faecal calprotectin.[19] Two excluded studies were performed in European countries[18,19] and

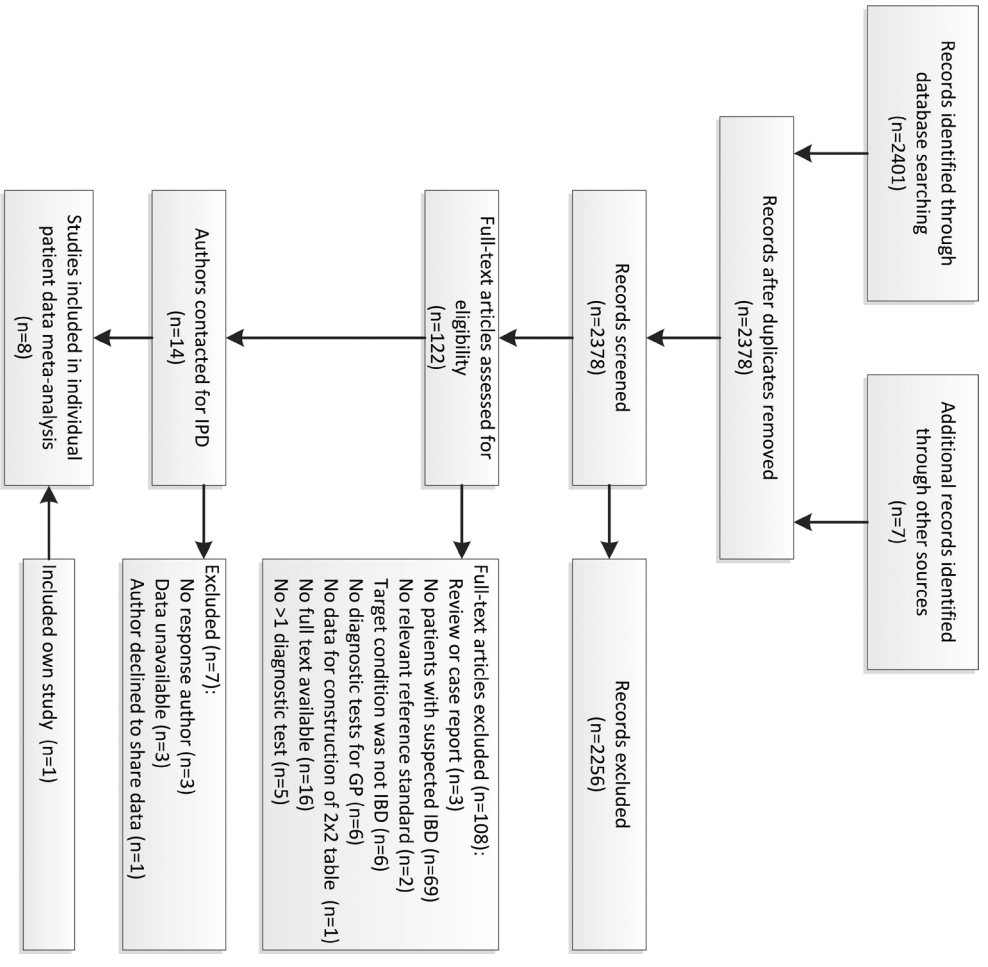


Figure 1. Flow diagram summarizing study identification and selection.

➤ **Table 1: Study and patient characteristics of eight included studies providing individual patient data.**

Note: referred moderate risk: children referred by their general practitioner (either general practitioner or paediatrician) to a paediatrician or paediatric gastroenterologist for diagnostic work-up; referred high risk: children referred by a paediatrician to a paediatric gastroenterologist and endoscopy; CRP (mg/l); ESR (mm/h); Platelets (x10^9/l); Hb (g/l) except for Perninow Hb (g/dl) and van de Vijver and Holtman Hb (mmol/l); Albumin (g/l); FCal (µg/g). Tsampalieros: abdominal pain and rectal bleeding were recorded for Crohn's disease (n=156) only. van de Vijver: abdominal pain included abdominal pain and diarrhoea. "–": variable missing. Abbreviations: CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, FCal: faecal calprotectin, Hb: haemoglobin, IBD: inflammatory bowel disease.

	Datasets															
	Fagerberg '05		*Henderson '12*		*Holtman '16*		*Leach '07*		*Perninow '09*		*Sidler '08*		*Tsampalieros '11*		*Van de Vijver'12*	
Study characteristics																
Country	Sweden		Scotland		The Netherlands		Australia		Norway		Australia		Canada		The Netherlands	
Design	Prospective cohort		Case-control		Prospective cohort		Nested case-control		Prospective cohort		Prospective cohort		Case-control		Prospective cohort	
Setting	Referred children High risk		Referred children High risk		Referred children Moderate risk		Referred children High risk		Referred children Moderate/high risk		Referred children High risk		Referred children High risk		Referred children Moderate/high risk	
Patient characteristics																
	IBD	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**
Number of patients	20	16	91	99	17	72	39	33	62	38	31	30	258	197	42	75
Age range (years)	6-17	6-17	3-18	1-16	8-17	4-17	5-14	5-12	2-18	1-18	2-16	2-16	1-17	2-17	6-18	6-18
Symptoms (n (%))																
Abdominal pain	15/19 (79)	13/15 (87)	-	-	15 (88)	42 (58)	-	-	50 (81)	32 (84)	17 (55)	21 (70)	123/156 (79)	-	29 (69)	67 (89)
Diarrhoea	18 (90)	11/14 (79)	-	-	16 (94)	45 (63)	-	-	42 (68)	19 (50)	18 (58)	7 (23)	-	-	8 (19)	2 (3)
Rectal bleeding	16 (80)	9 (56)	-	-	8 (47)	18 (25)	-	-	32 (52)	10/37 (27)	9 (29)	5 (17)	90/156 (58)	-	26 (62)	32 (43)
Weight loss	10/16 (63)	7/12 (58)	-	-	12 (71)	11 (15)	-	-	-	-	9 (29)	10 (33)	-	-	23 (55)	35 (47)
Blood markers (median (IQR))																
CRP	6.9 (7-10)	6.9 (7-7)	11 (3-35)	1 (1-3)	12 (3-22)	1 (0.3-3)	11 (1-43)	1 (1-2)	7.5 (6-28)	7 (5-7)	8 (3-48)	1 (1-1.5)	15.4 (4.4-45.5)	0 (0-2.5)	16 (5-34)	5 (5-5)
ESR	12 (5-27)	4 (2-8)	30 17-50)	6 (2-11)	25 (10-32)	6 (4-10)	28 (14-40)	3 (2-7)	22 (9-34)	5 (3-11)	27 (14-43)	5 (4-7)	34 (20-48)	12 (6.5-18.5)	26 (17-51)	7.5 (4-17)
Platelets	369 (314-450)	295 (273-300)	415 (327-505)	306 (239-343)	409 (257-448)	292 (251-332)	414 (331-534)	301 (246-341)	369 (300-427)	268 (223-325)	444 (323-529)	296 (254-340)	459 (347-562)	-	-	-
Hb	117.5 (106-128)	134 (126-145)	115 (100-123)	129 (120-135)	7.3 (7-8)	8.1 (8-9)	-	-	11.5 (10-13)	12.6 (12-13)	-	-	114 (103-124)	-	7 (6-8)	8 (8-8)
Albumin	36 (33-40)	43.5 (40-46)	38 (34-41)	46 (43-47)	-	-	33 (29-38)	40 (39-41)	38 (33-42)	41 (37-43)	33 (27-35)	40 (38-42)	36 (32-41)	-	-	-
Faecal marker (median (IQR))																
FCal	389 (219-713)	17.5 (8-28)	1265 (714-2035)	65 (20-235)	711 (470-824)	20 (20-30)	-	-	1183 (359-2000)	33 (14-113)	1265 (575-2517)	31 (20-67)	-	-	1408 (1067-1800)	40 (40-100)

five studies in North America.[20-24] The test characteristics of the laboratory markers of the available and excluded studies were comparable,[3] except for one excluded study that showed to be an outlier for CRP and platelet count.[22]

CHARACTERISTICS AND QUALITY OF INCLUDED STUDIES

We were able to obtain the IPD from eight studies with a total of 1120 paediatric patients of whom 560 had IBD. Study and patient characteristics of included studies are given in Table 1. The mean prevalence of IBD in the five cohort studies was 43% (range 19%-62%).[10,25-28] Five of the eight included studies were performed in European countries[10,25-27,29], two in Australia[28,30], and one in North America.[31] All studies were performed in referred children or adolescents (hospital setting); three used a case-control design in symptomatic paediatric patients.[29-31]

Table 2. Summary of the methodological assessment of eight included studies providing individual patient data.

	RISK OF BIAS				APPLICABILITY CONCERNS		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Cohort studies							
Fagerberg et al, 2005[25]	High	Unclear	Unclear	High	Low	Low	Low
Holtman et al, 2016[10]	Low	Low	Low	High	High	Low	Unclear
Perminow et al, 2009[26]	Unclear	Unclear	High	High	Low	Low	Low
Sidler et al, 2008[28]	Unclear	High	Low	Unclear	Low	Low	Low
Van de Vijver et al, 2012[27]	Unclear	Low	High	High	Low	Low	Low
Case-control studies							
Henderson et al, 2012[29]	High	Low	High	High	Low	Low	Low
Leach et al, 2007[30]	High	Unclear	High	Low	Low	Low	Low
Tsampalieros et al, 2011[31]	High	Unclear	Unclear	Unclear	Low	Low	Low

High risk of bias in domain patient selection if the study had no consecutive or random sample of patients, case-control design,[29-31] or inappropriate exclusions.[25,30,31] High risk of bias in domain index test if the index results were interpreted with knowledge of the reference standard, or if the threshold was not pre-specified.[28] High risk of bias in domain reference standard if the endoscopy did not include an ileum intubation,[29,30] the follow-up was less than 12 months,[27] or if the reference standard results were interpreted with knowledge of the index test.[26,29] High risk of bias in the domain flow and timing if the time period between index test and reference standard was more than 1 month,[27] not all patients receive a reference standard, not all patients receive the same reference standard,[10,27] or not all patients were included in the analysis.[25,26] High applicability concerns for the domain patient selection if not all children were at risk for IBD.[10]

The QUADAS-2 was scored on the published information. Fagerberg[25] published the study before the QUADAS-2 and mentioned to us that the patients were included consecutively in a prospective cohort study, the index results were not interpreted with knowledge of the reference standard and the reference standard were not interpreted with knowledge of the index test.

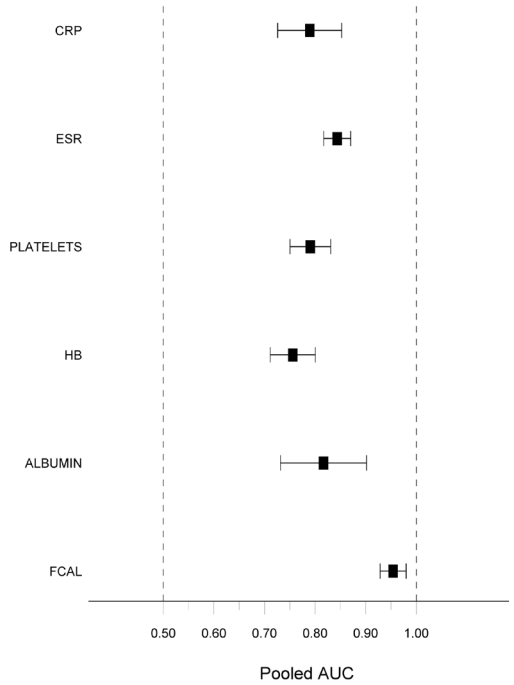


Figure 2. Forest plot showing pooled AUC with 95% CI. Abbreviations: AUC: area under the curve, CI, confidence interval, CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, FCal: faecal calprotectin, HB: haemoglobin.

Quality assessment of all included studies identified risk of bias in ≥1 domain. We had applicability concerns for patient selection in one study.[10] Table 2 presents the full QUADAS-2. Appendix 2 presents the systematically missing and sporadically missing values. The sporadically missing values were imputed.

DISCRIMINATION OF MARKERS

The pooled AUC of ESR (8 studies), albumin (5 studies), CRP (8 studies), platelets (6 studies), Hb (5 studies), and faecal calprotectin (6 studies) were 0.84 (0.82-0.87), 0.82 (0.73-0.90), 0.79 (0.73-0.85), 0.79 (0.75-0.83), 0.76 (0.71-0.80), and 0.95 (0.93-0.98), respectively (Figure 2; Appendix 3). The AUC of all markers, except for haemoglobin, were heterogeneous across studies.

ADDED VALUE OF MARKERS

In two studies the basic model could not be fitted, because of one or more of the key symptoms was systematically missing (Appendix 2).[29,30] The AUC of the basic model ranged from 0.65 to 0.77, and the pooled AUC of the basic model was 0.70 (0.65-0.75). Adding ESR (5 studies), platelets (4 studies), Hb (4 studies), albumin (3 studies), and CRP (5 studies) to the basic model of symptoms resulted in a pooled delta AUC of 0.16 (0.11-0.21), 0.13 (0.08-0.19), 0.13 (0.08-0.19), 0.13 (0.04-0.21), and 0.08 (0.04-0.11), respectively (Figure 3; Appendix 4). The improvement in AUC when adding faecal calprotectin to the basic model ranged from 0.21 to 0.29, and was statistically significant in all datasets. The pooled delta AUC of faecal calprotectin was 0.26 (0.21-0.31) (Figure 3; Appendix 4).

IMPROVEMENT IN DIAGNOSTIC RISK CLASSIFICATION

The reclassification table of 100 hypothetical paediatric patients with IBD prevalence of 43% illustrates that adding the best marker (faecal calprotectin) to the basic model of symptoms leads to a decrease in the intermediate risk group from 55 to 6 paediatric patients (Table 3).

The proportion of paediatric patients without IBD correctly classified as low risk of IBD increased from 33% to 91%. The proportion of IBD cases in the low risk group decreased when faecal calprotectin was added to symptoms (27% vs 7%) and increased in the high risk group (74% vs 95%).

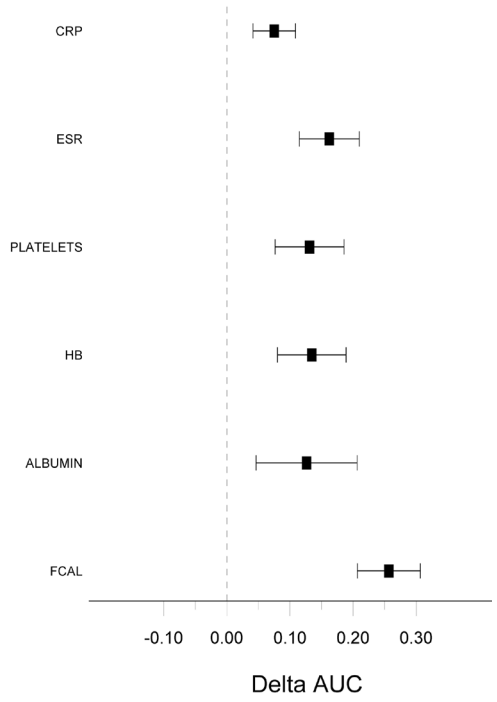


Figure 3. Forest plot showing the pooled improvement in AUC when adding markers to the basic model with 95% CI.

Note: a delta AUC value >0 implies an added discriminative value of the laboratory test, and a value ≤0 implies no added discriminative value. Abbreviations: AUC: area under the curve, CI, confidence interval, CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, FCal: faecal calprotectin, HB: haemoglobin.

DISCUSSION

This IPD meta-analysis, including 1120 referred paediatric patients with symptoms suggestive of IBD demonstrated that all laboratory markers (ESR, CRP, platelets, Hb, albumin and faecal calprotectin) as a single test improved the discrimination between patients with and without IBD when added to a model with symptoms alone. The addition of faecal calprotectin to symptoms improved the AUC more than any of the individual blood markers. Moreover, faecal calprotectin added to symptoms improved the diagnostic risk classification by decreasing the number of paediatric patients in the intermediate risk group from 55% to 6%. The paediatric patients were more often correctly classified in the low and high risk group after adding faecal calprotectin to the diagnostic process.

The basic model in different datasets performed poor to fair (AUC varied between 0.65-0.77). Of note we have to consider that the performance of discrimination of the basic model might have been better when more alarm symptoms would have been included in the model. This was not possible, since the included studies did often not record involuntary weight loss, growth failure, peri-anal lesions, family history of IBD, or extra-intestinal symptoms. We found that in referred symptomatic paediatric patients, all laboratory markers did add significant discriminative value to symptoms alone and hence are potentially of value in the triage for endoscopy. Clinical relevance, however, depends on treatment thresholds and the trade-off between the utility of a missed (or delayed) diagnosis of IBD and an unnecessary endoscopy under full anaesthesia. Guidelines suggest performing blood tests in paediatric patients with symptoms suggestive for IBD.[1,8] Because blood markers as haemoglobin and albumin also may have consequences for treatment choices this recommendation should not be abandoned. However, for the triage of paediatric patients for endoscopy, faecal calprotectin showed highest discriminative performance and should be recommended for this purpose. Especially, since a normal faecal calprotectin (<50 μg/g) can rule out an IBD diagnosis which is important

Table 3: Reclassification table providing the improved diagnostic risk classification after adding faecal calprotectin to symptoms in a hypothetical cohort of 100 children with an inflammatory bowel disease prevalence of 43%.

	Predicted risk of IBD	Observed IBD		
		Yes	No	Total
Basic model	Low <35%	7	19	26
	Intermediate 35%-60%	22	33	55
	High >60%	14	5	19
	Total risk group	43	57	100
Basic model + faecal calprotectin	Low <35%	4	52	56
	Intermediate 35%-60%	3	3	6
	High >60%	36	2	38
	Total risk group	43	57	100

The numbers in the table are based on the median prevalence of IBD of 43% across cohort studies and pooled sensitivities and specificities of the basic and the basic + faecal calprotectin model at low (35%) and high (60%) predicted probabilities of IBD. The pooled sensitivities for the basic and basic + faecal at low predicted probabilities were 0.84 and 0.91, respectively. The pooled specificities were 0.33 and 0.92. At high predicted probabilities the pooled sensitivities were 0.33 and 0.84, and pooled specificities were 0.92 and 0.96.

in the triage for endoscopy.[4,6] Normal blood tests do not rule out an IBD diagnosis.[3,32]

The results are applicable to clinicians who evaluate referred paediatric patients for symptoms suggestive for IBD. In only one study a quarter of the patients were initially assessed in primary care and referred to specialist care for further diagnostic work-up.[10] More studies in primary care are needed to evaluate the added value of markers in this setting.

COMPARISON WITH LITERATURE

This is the first meta-analysis using IPD to investigate the added value of commonly used laboratory markers for diagnosing IBD. However, another IPD meta-analysis concerning faecal calprotectin in referred paediatric patients with suspected IBD developed an individual risk prediction rule for IBD.[7] The prediction rule was based on faecal calprotectin value and the age of the child. The AUC of the prediction model was 0.92 (0.89-0.94). In daily practice, signs and symptoms are used before testing with blood markers or faecal calprotectin. Therefore, it is important to ascertain the incremental value of signs and symptoms alongside laboratory testing. In the current IPD meta-analysis we evaluated the most commonly used laboratory markers and provided insight into which tests are appropriate for triage for endoscopy.

Degraeuwe et al.[7] also found in their IPD meta-analysis that the AUC of testing with faecal calprotectin was 0.94 (0.92-0.95). In the current IPD meta-analysis the AUC of faecal calprotectin was comparable, even though we included different studies. Four studies included in the earlier IPD were not included in the current IPD, because two included only faecal

calprotectin testing,[33,34] one study included paediatric patients with known IBD,[35] and the authors of one study did not respond to our efforts to contact them.[19] In our IPD meta-analysis, we included two additional studies,[10,26] one of which was published after the earlier IPD.[10]

STRENGTHS AND LIMITATIONS

Of the 15 eligible studies, we were able to obtain datasets from eight studies. Variation in definitions of symptoms might explain heterogeneity and the variance between available and excluded studies. Because the test characteristics of the laboratory markers of the available and excluded studies were comparable, we expect that the excluded studies will not have a large impact on the results.

The median and AUC of some laboratory tests varied considerably between the included studies. These heterogeneous results might be explained by the different assays that were used for the laboratory tests. Moreover, the AUC may vary due to different designs (cohort or case control) and the number and choice of the reference standards (endoscopy or follow-up). However, the delta AUCs were more homogeneous than the AUCs. We chose a two-step approach, because this is a transparent method that takes into account the hierarchical nature of the data, which means that patients and procedures from one study are more consistent and similar to each other than across different studies.

Due to the absence of the registration of symptoms in three datasets,[29-31] it was not possible to evaluate the added value of the markers in these datasets. We did not ask the authors to retrospectively review the symptoms in the medical files, since this would make the information less reliable. Another limitation is that the number of patients in the included studies was small. Too many predictors for a low number of patients in the studies may cause perfect discrimination. The AUC of faecal calprotectin was very high, which might be an overestimation. Due to the high AUC of symptoms and faecal calprotectin there is a small chance that blood markers could have had added value. However, the number of paediatric patients in the included studies of this IPD meta-analysis was too small to evaluate the added value of blood markers to symptoms and faecal calprotectin. A much larger study with more patients with and without IBD is needed to develop a prediction model for IBD based on patient characteristics, single alarm symptoms, blood markers, and faecal calprotectin.

Since the AUC is an overall measure of discrimination and gives no insight to clinical interpretation, we provided a reclassification table of the best marker. In clinical practice, a threshold is used to classify patients into a low risk group to rule out IBD and into a high risk group to confirm IBD. We assume that when referred patients are classified into the low risk group (probability <35) the paediatrician decides not to perform an endoscopy, while patients in the high risk group (probability >60) are considered likely to have IBD and require an endoscopy to determine the diagnosis. This assumption and the thresholds of the risk groups may be debated. Other thresholds to define the three risk groups could change the reclassifications. However, 35 and 60 are reasonable thresholds in specialist care, because studies show that paediatric patients with a probability around 35% are referred to the paediatric gastroenterologist and paediatric patients with a probability around 60% received an endoscopy.[27,36] For the clinician, the intermediate risk group is the most challenging, because uncertainty about appropriate management is highest.

CONCLUSIONS

In referred paediatric patients faecal calprotectin added the most diagnostic value to symptoms compared to commonly used blood markers. Addition of faecal calprotectin to the diagnostic work-up of referred paediatric patients with symptoms suggestive of IBD considerably decreased the number of paediatric patients in the intermediate risk for IBD group.

3

REFERENCES

1. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr*. 2014;58(6):795-806.
2. Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM. Quantifying the added value of a diagnostic test or marker. *Clin Chem*. 2012;58(10):1408-1417.
3. Holtman GA, Lisman-van Leeuwen Y, Reitsma JB, Berger MY. Noninvasive tests for inflammatory bowel disease: A meta-analysis. *Pediatrics*. 2016;137(1):1-11.
4. van Rheenen PF, Van de Vijver E, Fidler V. Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: Diagnostic meta-analysis. *BMJ*. 2010;341:c3369.
5. Waugh N, Cummins E, Royle P, et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: Systematic review and economic evaluation. *Health Technol Assess*. 2013;17(55):xv-xix, 1-211.
6. Henderson P, Anderson NH, Wilson DC. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease: A systematic review and meta-analysis. *Am J Gastroenterol*. 2014;109(5):637-645.
7. Degraeuwe PL, Beld MP, Ashorn M, et al. Faecal calprotectin in suspected paediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2015;60(3):339-346.
8. de Ridder L, Rings EH, Escher JC, CBO-werkgroep 'IBD bij kinderen'. Guideline 'diagnosis and treatment of inflammatory bowel disease in children'. *Ned Tijdschr Geneeskd*. 2010;154:A1898.
9. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
10. Holtman GA, Lisman-van Leeuwen Y, Kollen BJ, et al. Challenges in diagnostic accuracy studies in primary care: The fecal calprotectin example. *BMC Fam Pract*. 2013;14:179-2296-14-179.
11. Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. in: Higgins JPT, green S (editors). cochrane handbook for systematic reviews of interventions version 5.1.0 (updated march 2011). the cochrane collaboration, 2011. available from www.cochrane-handbook.org).
12. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*. 1988:837-845.
13. Pepe M, Longton G, Janes H. Estimation and comparison of receiver operating characteristic curves. *Stata J*. 2009;9(1):1.
14. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.
15. van der Heijden, Geert JMG, Donders ART, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. *J Clin Epidemiol*. 2006;59(10):1102-1109.
16. Moons KG, Donders RA, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59(10):1092-1101.
17. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines. *BMC Med Res Methodol*. 2009;9:57-2288-9-57.
18. Cabrera-Abreu JC, Davies P, Matek Z, Murphy MS. Performance of blood tests in diagnosis of inflammatory bowel disease in a specialist clinic. *Arch Dis Child*. 2004;89(1):69-71.
19. Canani RB, de Horatio LT, Terrin G, et al. Combined use of noninvasive tests is useful in the initial diagnostic approach to a child with suspected inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2006;42(1):9-15.
20. Khan K, Schwarzenberg SJ, Sharp H, Greenwood D, Weisdorf-Schindele S. Role of serology and routine laboratory tests in childhood inflammatory bowel disease. *Inflamm Bowel Dis*. 2002;8(5):325-329.
21. Dubinsky MC, Ofman JJ, Urman M, Targan SR, Seidman EG. Clinical utility of serodiagnostic testing in suspected pediatric inflammatory bowel disease. *Am J Gastroenterol*. 2001;96(3):758-765.
22. Beattie RM, Walker-Smith JA, Murch SH. Indications for investigation of chronic gastrointestinal symptoms. *Arch Dis Child*. 1995;73(4):354-355.
23. Sabery N, Bass D. Use of serologic markers as a screening tool in inflammatory bowel disease compared with elevated erythrocyte sedimentation rate and anemia. *Pediatrics*. 2007;119(1):e193-9.
24. El-Chammas K, Majeskie A, Simpson P, Sood M, Miranda A. Red flags in children with chronic abdominal pain and crohn's disease-a single center experience. *J Pediatr*. 2013;162(4):783-787.
25. Fagerberg UL, Loof L, Myrdal U, Hansson LO, Finkel Y. Colorectal inflammation is well predicted by fecal calprotectin in children with gastrointestinal symptoms. *J Pediatr Gastroenterol Nutr*. 2005;40(4):450-455.
26. Perminow G, Brackmann S, Lyckander LG, et al. A characterization in childhood inflammatory bowel disease, a new population-based inception cohort from south-eastern norway, 2005-07, showing increased incidence in crohn's disease. *Scand J Gastroenterol*. 2009;44(4):446-456.
27. Van de Vijver E, Schreuder AB, Cnossen WR, Muller Kobold AC, van Rheenen PF, North Netherlands Pediatric IBD Consortium. Safely ruling out inflammatory bowel disease in children and teenagers without referral for endoscopy. *Arch Dis Child*. 2012;97(12):1014-1018.
28. Sidler MA, Leach ST, Day AS. Fecal S100A12 and fecal calprotectin as noninvasive markers for inflammatory bowel disease in children. *Inflamm Bowel Dis*. 2008;14(3):359-366.
29. Henderson P, Casey A, Lawrence SJ, et al. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease. *Am J Gastroenterol*. 2012;107(6): 941-949.
30. Leach ST, Yang Z, Messina I, et al. Serum and mucosal S100 proteins, calprotectin (S100A8/S100A9) and S100A12, are elevated at diagnosis in children with inflammatory bowel disease. *Scand J Gastroenterol*. 2007;42(11):1321-1331.
31. Tsampalieros A, Griffiths AM, Barrowman N, Mack DR. Use of C-reactive protein in children with newly diagnosed inflammatory bowel disease. *J Pediatr*. 2011;159(2):340-342.
32. Mack DR, Langton C, Markowitz J, et al. Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics*. 2007;119(6):1113-1119.
33. Ashorn S, Honkanen T, Kolho KL, et al. Fecal calprotectin levels and serological responses to microbial antigens among children and adolescents with inflammatory bowel disease. *Inflamm Bowel Dis*. 2009;15(2):199-205.
34. Diamanti A, Panetta F, Basso MS, et al. Diagnostic work-up of inflammatory bowel disease in children: The role of calprotectin assay. *Inflamm Bowel Dis*. 2010;16(11):1926-1930.
35. Kolho KL, Raivio T, Lindahl H, Savilahti E. Fecal calprotectin remains high during glucocorticoid therapy in children with inflammatory bowel disease. *Scand J Gastroenterol*. 2006;41(6):720-725.
36. Heida A, Holtman GA, Lisman-van Leeuwen Y, Berger MY, van Rheenen PF. Avoid endoscopy in children with suspected inflammatory bowel disease who have normal calprotectin levels. *J Pediatr Gastroenterol Nutr*. 2016;62(1):47-49.

3

Appendix 1. Full text of search strategy

SEARCH STRATEGY PUBMED

Target condition IBD
("Inflammatory bowel diseases"[MeSH] OR inflammatory bowel disease*[tw] OR IBD[tw] OR Colitis [tw] OR Crohn [tw])

Child
("Child"[Mesh] OR "Adolescent"[Mesh] OR "Infant"[Mesh]OR child*[tw] OR infant*[tw] OR adolescent*[tw] OR pediatric[tw] OR paediatric[tw] OR teenager*[tw])

Diagnostic accuracy
("Sensitivity and Specificity"[Mesh] OR specificit*[tw] OR false negative[tw] OR accura*[tw] OR sensitiv*[tw] OR "reproducibility of results"[MeSH])

SEARCH STRATEGY EMBASE

Target condition IBD
('enteritis'/de OR 'colitis'/exp OR 'inflammatory bowel disease':de,ab,ti OR IBD:de,ab,ti OR colitis:de,ab,ti OR crohn:de,ab,ti)

Child
('child'/exp OR 'newborn'/exp OR 'adolescent'/exp OR child*:de,ab,ti OR infant*:de,ab,ti OR adolescent*:de,ab,ti OR pediatric:de,ab,ti OR paediatric:de,ab,ti OR teenager*:de,ab,ti)

Diagnostic accuracy
('sensitivity and specificity'/exp OR 'predictive value'/exp OR 'receiver operating characteristic'/exp OR sensitiv*:de,ab,ti OR specificit*:de,ab,ti OR accura*:de,ab,ti OR 'false negative':de,ab,ti OR 'reproducibility'/exp) NOT [medline]/lim AND [embase]/lim

Appendix 2. The available information of symptoms, blood markers, and faecal calprotectin.

	Fagerberg '05		*Henderson '12*		*Holtman '16*		*Leach '07*		*Perminow '09*		*Sidler '08*		*Tsampalieros'11*		*Van de Vijver'12*	
	IBD	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**	**IBD**	**No IBD**
Number of patients	20	16	91	99	17	72	39	33	62	38	31	30	258	197	42	75
	missing		**missing**		**missing**		**missing**		**missing**		**missing**		**missing**		**missing**	
Symptoms																
Abdominal pain	5%	6%	**100%**	**100%**	0%	0%	**100%**	**100%**	0%	0%	0%	0%	40%	**100%**	0%	0%
Diarrhoea	0%	13%	**100%**	**100%**	0%	0%	**100%**	**100%**	0%	0%	0%	0%	**100%**	**100%**	0%	0%
Rectal bleeding	0%	0%	**100%**	**100%**	0%	0%	**100%**	**100%**	0%	3%	0%	0%	40%	**100%**	0%	0%
Weight loss	20%	25%	**100%**	**100%**	0%	0%	**100%**	**100%**	**100%**	**100%**	0%	0%	**100%**	**100%**	0%	0%
Blood markers																
CRP	0%	0%	3%	9%	12%	18%	28%	18%	3%	3%	0%	3%	0%	0%	2%	12%
ESR	0%	0%	4%	13%	0%	10%	28%	12%	13%	24%	0%	3%	1%	2%	2%	9%
Platelets	0%	0%	0%	3%	6%	6%	28%	12%	2%	0%	0%	0%	0%	**100%**	**100%**	**100%**
Hb	0%	0%	0%	2%	6%	6%	**100%**	**100%**	0%	0%	**100%**	**100%**	0%	**100%**	2%	4%
Albumin	0%	0%	0%	13%	**100%**	**100%**	28%	21%	5%	5%	0%	0%	0%	**100%**	**100%**	**100%**
Faecal marker																
FCal	0%	0%	0%	0%	6%	4%	**100%**	**100%**	5%	8%	3%	0%	**100%**	**100%**	0%	0%

Note: Henderson 2012, Leach 2007, Tsampalieros 2011 did not collected information on the symptoms of all children. Abbreviations: CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, FCal: faecal calprotectin, Hb: haemoglobin, IBD: inflammatory bowel disease.

Appendix 3. The area under the curve of the individual laboratory markers as observed in eight different datasets together with a random effects pooled estimate.

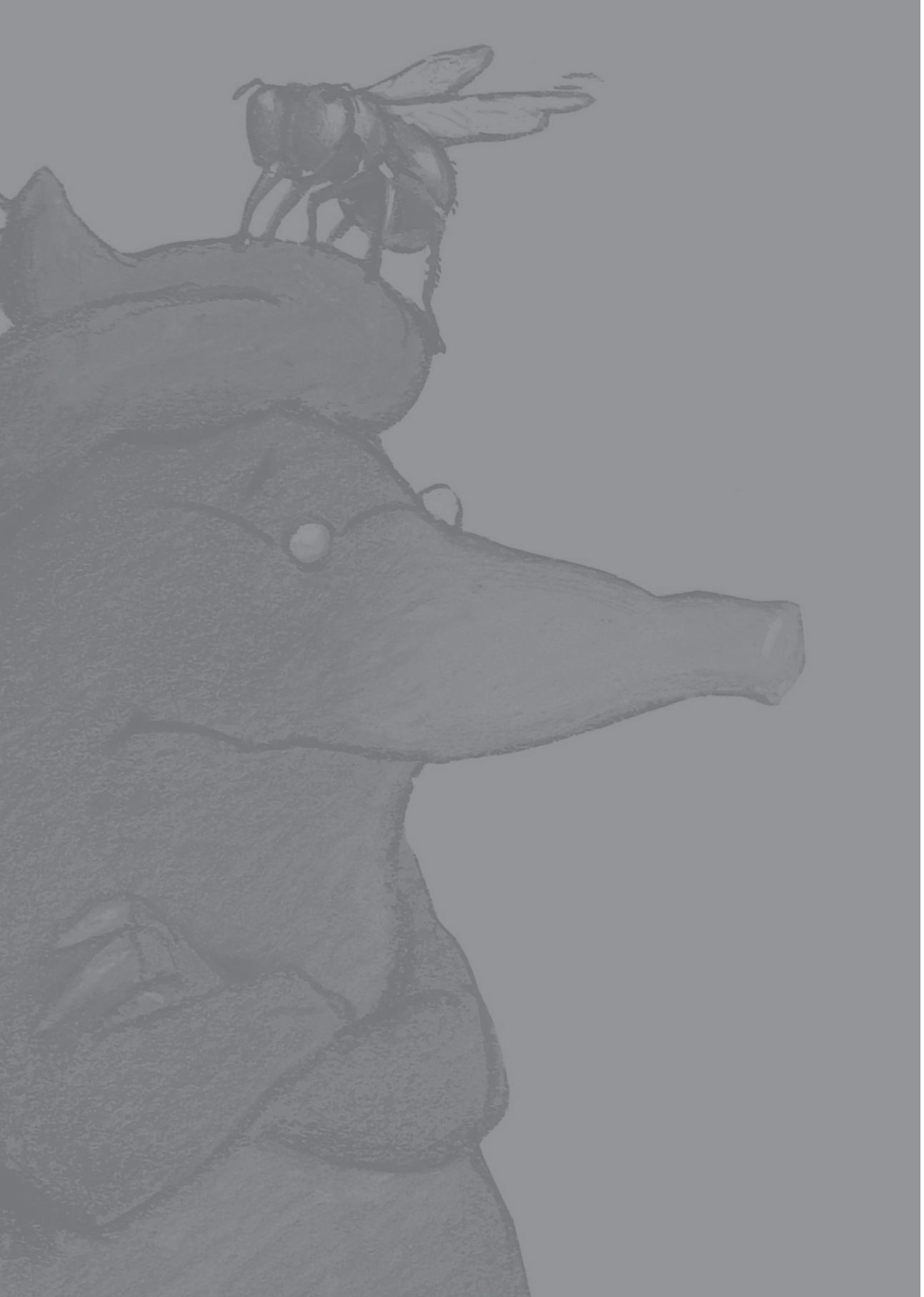
	AUC CRP	AUC ESR	AUC platelets	AUC Hb	AUC albumin	AUC FCal
Fagerberg '05	0.63 (0.45-0.81)	0.76 (0.60-0.92)	0.78 (0.62-0.94)	0.77 (0.60-0.94)	0.83 (0.69-0.97)	0.97 (0.92-1.00)
Henderson '12	0.83 (0.77-0.89)	0.84 (0.77-0.90)	0.79 (0.72-0.85)	0.76 (0.69-0.83)	0.90 (0.85-0.94)	0.93 (0.90-0.97)
Holtman '16	0.80 (0.69-0.90)	0.80 (0.66-0.94)	0.72 (0.56-0.88)	0.78 (0.65-0.92)	-	0.98 (0.94-1.00)
Leach '07	0.82 (0.72-0.92)	0.88 (0.81-0.96)	0.81 (0.70-0.91)	-	0.80 (0.69-0.91)	-
Perminow '09	0.61 (0.50-0.73)	0.77 (0.68-0.87)	0.79 (0.70-0.89)	0.70 (0.59-0.80)	0.67 (0.56-0.78)	0.87 (0.80-0.95)
Sidler '08	0.89 (0.80-0.98)	0.91 (0.83-1.00)	0.81 (0.71-0.92)	-	0.86 (0.75-0.96)	0.98 (0.96-1.00)
Tsampalieros '11	0.88 (0.85-0.91)	0.83 (0.80-0.87)	-	-	-	-
Van de Vijver '12	0.74 (0.64-0.84)	0.86 (0.80-0.93)	-	0.79 (0.69-0.88)	-	0.95 (0.91-0.99)
Pooled	**0.79 (0.73-0.85)**	**0.84 (0.82-0.87)**	**0.79 (0.75-0.83)**	**0.76 (0.71-0.80)**	**0.82 (0.73-0.90)**	**0.95 (0.93-0.98)**

Abbreviations: CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, FCal: faecal calprotectin, Hb: haemoglobin AUC: area under the receiver operating characteristic curve.

Appendix 4. The improvement in area under the curve when adding markers to the basic model and random effects pooled estimate.

	delta AUC CRP	delta AUC ESR	delta AUC platelets	delta AUC Hb	delta AUC albumin	delta AUC FCal
Fagerberg '05	0.12 (-0.02-0.26)	0.10 (-0.05-0.26)	0.17 (0.01-0.34)	0.10 (-0.07-0.28)	0.17 (0.01-0.33)	0.29 (0.13-0.45)
Holtman '16	0.15 (0.06-0.24)	0.14 (0.05-0.23)	0.09 (0.00-0.18)	0.15 (0.05-0.25)	-	0.23 (0.13-0.32)
Perminow '09	0.04 (-0.02-0.10)	0.15 (0.04-0.25)	0.17 (0.06-0.28)	0.10 (0.00-0.19)	0.07 (-0.01-0.16)	0.21 (0.10-0.32)
Sidler '08	0.05 (0.00-0.11)	0.18 (0.06-0.30)	0.14 (0.03-0.25)	-	0.19 (0.06-0.32)	0.29 (0.16-0.42)
Van de Vijver '12	0.08 (0.03-0.13)	0.21 (0.12-0.30)	-	0.17 (0.07-0.27)	-	0.29 (0.20-0.39)
Pooled	**0.08 (0.04-0.11)**	**0.16 (0.11-0.21)**	**0.13 (0.08-0.19)**	**0.13 (0.08-0.19)**	**0.13 (0.04-0.21)**	**0.26 (0.21-0.31)**

Note: A delta AUC value > 0 implies an added discriminative value of the laboratory test, and a value ≤ 0 implies no added discriminative value. Abbreviations: CRP: C-reactive protein, ESR: erythrocyte sedimentation rate, FCal: faecal calprotectin, Hb: haemoglobin AUC: area under the receiver operating characteristic curve.



CHAPTER 4

CHALLENGES IN DIAGNOSTIC ACCURACY STUDIES IN PRIMARY CARE: THE FAECAL CALPROTECTIN EXAMPLE

Gea A Holtman

Yvonne Lisman-van Leeuwen

Boudewijn J Kollen

Johanna C Escher

Angelika Kindermann

Patrick F van Rheenen

Marjolein Y Berger

BMC Fam Pract. 2014 14:179.

ABSTRACT

BACKGROUND

Low disease prevalence and lack of uniform reference standards in primary care induce methodological challenges for investigating the diagnostic accuracy of a test. We present a study design that copes with these methodological challenges and discuss the methodological implications of our choices, using a quality assessment tool for diagnostic accuracy studies (QUADAS-2).

DESIGN

The study investigates the diagnostic value of faecal calprotectin for detecting inflammatory bowel disease in children presenting with chronic gastrointestinal symptoms in primary care. It is a prospective cohort study including two cohorts of children: one cohort will be recruited in primary care and the other in secondary/tertiary care. Test results of faecal calprotectin will be compared to one of the two reference standards for inflammatory bowel disease: endoscopy with histopathological examination of mucosal biopsies or assessment of clinical symptoms at 1-year follow-up.

DISCUSSION

According to QUADAS-2 the use of two reference standards and the recruitment of patients in two populations may cause differential verification bias and spectrum bias, respectively. The clinical relevance of this potential bias and methods to adjust for this are presented. This study illustrates the importance of awareness of the different kinds of bias that result from choices in the design phase of a diagnostic study in a low prevalence setting. This approach is exemplary for other diagnostic research in primary care.

INTRODUCTION

In primary care, patients often present with non-specific symptoms and the incidence of severe illnesses is low. Differentiating between innocent symptoms and a rare, but serious organic disease is a diagnostic dilemma for the general practitioner (GP). Unnecessary referrals and diagnostic testing need to be balanced against the risk of missing a diagnosis and introduction of an unacceptable long diagnostic delay. In primary care, both the GP and the patient would greatly benefit from simple, non-invasive and specific triage tests. However, many of these tests are not validated in primary care.

An example of such a diagnostic dilemma are children presenting with chronic or recurrent gastrointestinal symptoms. This clinical picture is common, but few children will actually have inflammatory bowel disease (IBD), which includes Crohn's disease and ulcerative colitis. The incidence of non-specific abdominal pain in Dutch children is 2500/100,000 per year, while the incidence of IBD is 5.2/100,000 per year.[1,2] Clinical symptoms in children with IBD are often non-specific and show substantial overlap with functional gastrointestinal disorders.[3] In European secondary and tertiary care facilities the measurement of calprotectin in stool is used as an effective triage method for endoscopy, which is the reference standard for the diagnosis of IBD.[4] Calprotectin is a marker of inflammation that can be measured by using a simple non-invasive test,[5] but has never been evaluated in children in a primary care setting.[6-8] The different patient spectrum in primary care has consequences for the pre-test probability and test characteristics. Before calprotectin testing can be recommended to distinguish functional from organic gastrointestinal disorders at the primary care level, information is required on the predictive value of faecal calprotectin at the primary care level.

The preferred design to evaluate the diagnostic value of faecal calprotectin in children with chronic gastrointestinal symptoms would be a cross-sectional study. Such a design has two methodological challenges. Firstly, the design of a diagnostic study for rare diseases requires a large population in order to identify a sufficient number of children with IBD; the financial and logistic exercise involved makes such a study infeasible.[9] Secondly, the preferred reference standard to detect IBD is endoscopy;[10] but it is unethical to perform this invasive test in children with a low likelihood of organic gastrointestinal disease.

Here we present an example of a design that copes with these methodological challenges. The methodological implications of applied design choices are examined using an evidence-based quality assessment tool for diagnostic accuracy studies (QUADAS-2).[11]

DESIGN

DESIGN AND SETTING

The DOK (*Darm Onderzoek bij Kinderen*; translated as *bowel research in children*) study is a prospective cohort study with a follow-up period of one year, also known as a delayed type cross-sectional study[12]. The study consists of two prospective cohorts. We will recruit a primary care cohort of children presenting consecutively in primary care in the northern part of the Netherlands (primary care cohort). A second cohort consists of children that will

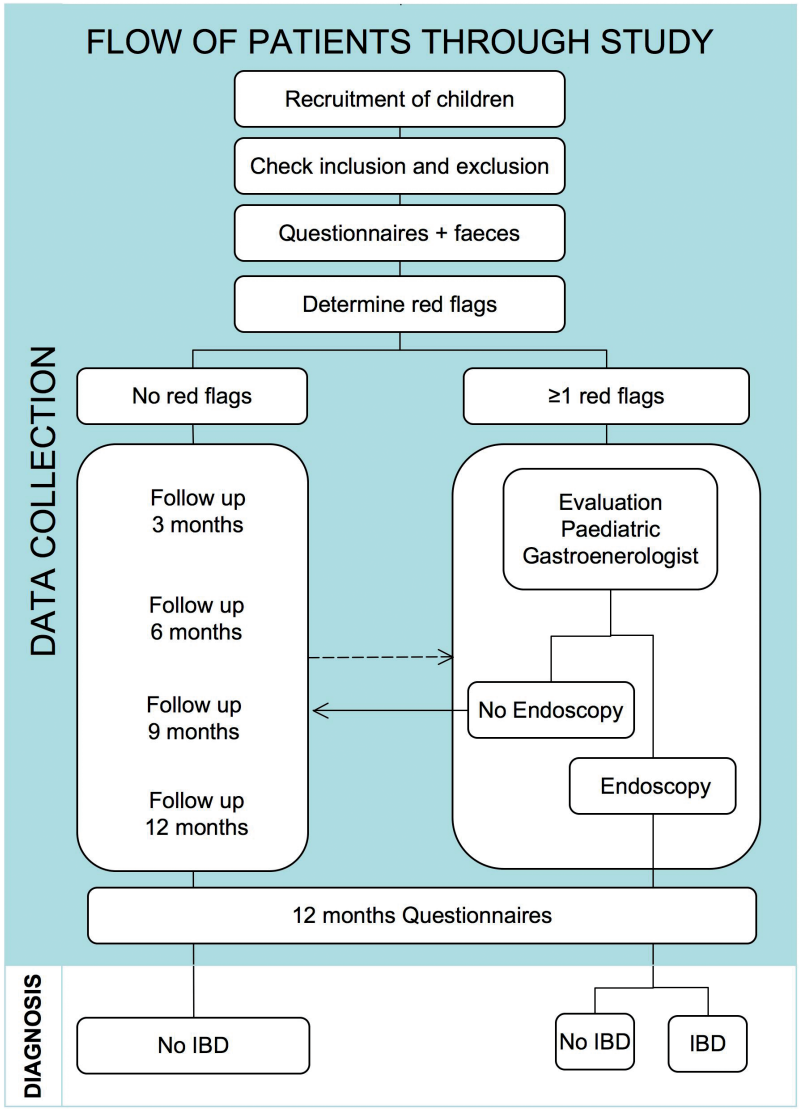


Figure 1. Flow chart of the DOK study.
The GP or paediatric gastroenterologist selects eligible children. At baseline inclusion, exclusion criteria and red flags are determined. The parents and child ≥10 years complete two questionnaires, i.e. a Questionnaire on Paediatric Gastrointestinal Symptoms (QPGS) and a symptoms questionnaire, in addition faeces (parasites and colon pathogens) are obtained. Children meeting ≥1 red flags are evaluated for eligibility for endoscopy by a paediatric gastroenterologist. Children without red flags receive a 1-year follow-up. The arrows indicate that the GP can refer a child during follow-up for endoscopic evaluation and the children who are not eligible for endoscopy receive a follow-up. After 1 year, information about diagnosis and clinical symptoms is collected based on the two above-mentioned questionnaires.

be referred to secondary and tertiary care facilities across the Netherlands (referred cohort). The index test is faecal calprotectin and the two reference standards for IBD are endoscopy with histopathological examination of mucosal biopsies, or assessment of clinical symptoms at 1-year follow-up (Figure 1).[4,13] The DOK study was approved by the Medical Ethics Review Committee of the University Medical Center Groningen. Written informed consent will be obtained from the parents and from the child if aged ≥12 years. Inclusion started in June 2011.

STUDY POPULATION

Children aged 4-18 years presenting with chronic diarrhoea (≥2 weeks diarrhoea or ≥2 episodes of diarrhoea in the past 6 months) or recurrent abdominal pain (≥2 episodes of abdominal pain in the past 6 months) will be eligible for participation. Diarrhoea was defined as moderately to watery loose stools matching score 5, 6 or 7 of the Bristol Stool Form Scale.[14] One episode is defined as 3 days or more.

Exclusion criteria are: a previously established diagnosis of chronic organic gastrointestinal disease; a complete evaluation in the past 6 months for abdominal symptoms including endoscopy; chronic use of antibiotics, non-steroidal anti-inflammatory drugs (NSAIDs) or oral corticosteroids (defined as daily use during ≥3 months/year); faecal calprotectin test in the past 6 months; and difficulty in understanding questionnaires. The number of patients not participating due to the exclusion criteria or refusal are anonymously recorded, including the patient characteristics and, if available, the reason for non-participation.

MEASUREMENTS

Physical examination

The GP or paediatric gastroenterologist performs a structured physical examination and assesses extra-intestinal symptoms and peri-anal lesions according to the Dutch diagnostic guideline.[15] The participating GPs receive training on structured physical examination of children with symptoms suggestive of IBD.

Questionnaire on Paediatric Gastrointestinal Symptoms

The Dutch version of the Questionnaire on Paediatric Gastrointestinal Symptoms ROME III (QPGS-RIII)[16] is completed, by the patient or a parent at baseline and at 12 months follow-up. The QPGS-RIII consists of two reports, a parent report for children aged 4-18 years and a self-report for children aged ≥10 years. The questionnaire has been translated into Dutch. The English version of QPGS has good content validity and test-retest reliability.[17,18]

Blood and faecal tests

In the blood sample haemoglobin, erythrocyte sedimentation rate, C-reactive protein, platelet count and serology tests for celiac disease (IgA tissue transglutaminase antibodies) are measured. Faeces is tested for colon pathogens (*Salmonella enterica, Campylobacter jejuni, Shigella spp/EIEC, STEC*) and parasites (*Giardia lamblia, Cryptosporidium spp, Dientamoeba fragilis, Entamoebe histolytica*) with the real-time multiplex PCRs.[19] Blood and faeces tests are performed at local certified laboratories. If a child is using NSAIDs, antibiotics or oral corticosteroids for short-term use (<3 months), the collection and testing of faeces is postponed until the end of that treatment.

Faecal calprotectin
After baseline assessments the patients send the faeces sample by pre-stamped return envelope to the laboratory where the samples are stored at -80°C. At the end of the data collection period the samples are defrosted before analysis. Faecal calprotectin is measured by means of a commercially available quantitative enzyme-linked immunosorbent assay (ELISA).[20,21] In accordance with the manufacturer's guidelines, values above 50 µg/g faeces are regarded as positive.

Red flags
In all children red flags of IBD will be searched for using a structured evaluation form (Table 1). Children who fulfil the inclusion criteria and have ≥1 red flags are referred to a paediatric gastroenterologist who will decide whether the child requires endoscopic examination.[15] This decision will be based on the medical history, physical examination and blood testing. Children without red flags, or those who are not eligible for endoscopy will be followed for one year.

Table 1. Definitions of red flag symptoms for inflammatory bowel disease.

Red flag symptom	Measurement	Positive
Growth failure	History and physical examination	Target height range > -1 SDS
Involuntary weight loss	History	Involuntary decrease in weight of > 1 kg
Rectal blood loss	History	Rectal blood loss with defecation without constipation according to ROME III criteria
Positive family history of IBD	History	Affected first-degree relative(s)
Extra-intestinal symptoms	Physical examination	Eyes (episcleritis, scleritis, uveitis), skin (erythema nodosum, pyoderma gangrenosum, psoriasis), mouth ulcers, finger clubbing, arthritis
Peri-anal lesions	Physical examination	Skin tags, haemorrhoids, fissures, fistulas, or abscess
Anaemia (Hb)	Local laboratory	4-12 years < 7.1 mmol/l, boy 12-18 years <8.1 mmol/l, girl 12-18 years <7.4 mmol/l [22]
CRP	Local laboratory	> 10 mg/l [23]
ESR	Local laboratory	> 20 mm/h [23]
Platelets	Local laboratory	> 450 x 10^9/l [24]

SDS = standard deviation score; Hb = haemoglobin; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate.

Endoscopy
Endoscopy is performed under full anaesthesia or deep sedation by an experienced paediatric gastroenterologist and entails esophagogastroduodenoscopy and ileocolonoscopy. Two biopsies of each intestinal segment are taken. The histopathological examination will be performed by an experienced gastrointestinal histopathologist. IBD is classified according to the Paris classification.[10]

Follow-up
Follow-up is done using a symptom questionnaire that was developed for the study in cooperation with paediatric gastroenterologists and GPs. This questionnaire will be completed by the parent or child (if aged ≥10 years) at 3, 6, 9 and 12 months follow-up. The GP will perform a structured physical examination to assess red flags in children with clinical symptoms at 12 months. Those with ≥1 red flags at 12 months will be referred to a paediatric gastroenterologist to determine a diagnosis.

Blinding
The paediatric gastroenterologists, pathologists, GPs and researchers will be blinded to the outcome of the faecal calprotectin test. The laboratory technician will be blinded for the clinical characteristics of the child and the result of endoscopy.

OUTCOME
IBD is confirmed when the endoscopic picture and the histopathological picture match. Absence of IBD is defined as a negative endoscopic and histopathological examination, or when there was no indication to perform endoscopy at all during the 12 months follow-up. Besides, all children without red flags at 12 months follow-up are considered not to have IBD.[13]

SAMPLE SIZE
Based on available literature we expect to find a specificity of 93% in the primary care cohort.[7,25-27] To estimate the specificity and a 95% confidence interval (CI) spanning 5%, we assume a maximum IBD incidence of 5 per 100 children with gastrointestinal complaints and a loss to follow-up of 10%, we will then need a sample size of 118 children in the primary care cohort. In a worst case scenario with a specificity of 75%, a sample size of 118 children will widen the 95% CI to 8%.[28]

Sensitivity was calculated in children with red flags (primary care cohort and referred cohort). Based on an expected sensitivity of 95% we need to include 73 children with IBD in order to estimate the sensitivity and a 95% CI spanning 5%.[7,26,27] With a IBD prevalence of 80% and a loss to follow-up of 10% we need to include 100 children with red flags. The prevalence of IBD is difficult to estimate; with a prevalence of 20% the spanning of the 95% CI of the sensitivity will widen to 10%.[28]

STATISTICAL ANALYSES
Specificity of faecal calprotectin for IBD in primary care will be calculated by dividing the number of negative faecal calprotectin tests by the total number of children without IBD

included in the primary care cohort. Sensitivity will be calculated by dividing the number of positive faecal calprotectin tests by the total number of children with IBD in children with red flags of both the primary care and referred cohort. The estimates of specificity and sensitivity will be reported as percentages with 95% CIs.

DISCUSSION

ASSESSING THE RISK OF BIAS

To address the risk of bias in our study design and the applicability of the results we applied the QUADAS-2 checklist[11] that includes four domains: patient selection, index test, reference standard and flow and timing (flow of patients through the study and timing of the index test and reference standard). Each domain was scored as low or high risk of bias, based on the answers to the signalling questions. If all answers concerning a domain are "yes", the risk of bias can be judged as low. If any signalling question is answered "no" the risk of bias can be judged as high. The first two domains were scored as low or high concerns regarding applicability. Two items were excluded because one item assessed heterogeneity between studies, which is only applicable in systematic reviews. The second item asked whether all patients are included in the analysis, which can only be assessed after completion of the study. The results of the QUADAS-2 assessment are shown in Table 2.

RISK OF BIAS

Problems with the reference standard

A perfect reference standard in a diagnostic accuracy study is said to fulfil three criteria: "1) The reference standard provides error-free classification of all subjects. 2) The same reference standard is used to verify all index results. 3) The index test and reference standard can be performed within a short interval to avoid changes in target condition status."[31]

Risk of bias in the DOK study is related to the choice of the reference test, which is not the same for all included patients (differential verification bias).[29] In addition, follow-up is not considered a reference standard for IBD in daily practice. This choice may lead to missed diagnoses and will influence the estimates of sensitivity and specificity. We chose a differential verification design, because it is unethical to perform endoscopy in children with a low likelihood of organic gastrointestinal disease. Therefore, children who have a low IBD risk receive a follow-up of one year, which is considered to be a suitable period.[13] On the opposite side, it might be possible that even more children will be identified because, using a 1-year follow-up, children with initially mild IBD can be detected when they have an aggravation of symptoms later in time. These children could have been missed when endoscopy was performed at initial presentation. Children in whom endoscopy was not indicated during the 1-year follow-up (either because they no longer have symptoms or because their red flags are not suggestive for IBD) are considered not to have IBD. The probability that we will miss a child with IBD is considered to be extremely low.[13] Adjustment for differential verification bias will be made, if possible, using a Bayesian approach.[29,30]

The patient flow of the DOK study could introduce bias. A delay of one month between

Table 2. Quality assessment of the DOK study design (QUADAS-2).

Signalling questions	Answer	Risk of Bias/ Applicability	Planned adjustment
Domain 1: Patient selection			
Risk of bias		**Low Risk**	
Is a consecutive sample of patients enrolled?	Yes		
Is a case-control design avoided?	Yes		
Does the study avoid inappropriate exclusions?	Yes		
Applicability		**High Concern**	- Magnitude will be evaluated
Are there concerns that the included patients and setting do not match the topic of our study (patients had symptoms suggestive of inflammatory bowel disease in primary care)?			
Domain 2: Index test			
Risk of bias		**Low Risk**	
Are the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, is it pre-specified?	Yes		
Applicability		**Low Concern**	
Are there concerns that the index test, its conduct, or interpretation differ from the topic of our study (faecal calprotectin was measured with ELISA)?			
Domain 3: Reference standard			
Risk of bias		**High Risk**	
Is the reference standard likely to correctly classify the target condition?	No		- Probably not clinically relevant - Adjustment in analysis [29,30]
Are the reference standard results interpreted without knowledge of the results of the index test?	Yes		
Domain 4: Flow and timing			
Risk of bias		**High Risk**	
Is there an appropriate interval between index test and reference standard?	No		- Represents care as usual - Repeated measurement index test before endoscopy
Do all patients receive a reference standard?	Yes		
Do all patients receive the same reference standard?	No		- Adjustment in analysis [29,30]

stool sample collection and reference standard is considered to be an appropriate time period. In children of the referred cohort the interval between faecal sampling and endoscopy will generally be less than one month. For referred children in the primary care cohort this interval is likely to exceed the period of one month. To investigate whether the concentration of faecal calprotectin accurately measures the same outcome as endoscopy, faeces will be collected again shortly before endoscopy. In children not referred to a secondary or tertiary care facility the period between faecal sampling and reference test will be one year. During this period the calprotectin concentration may change and, therefore, the initial test result will no longer be related to the outcome of endoscopy. This will underestimate sensitivity and specificity. Here we adopt a pragmatic approach. We want to establish whether faecal calprotectin can serve as a triage test in children who are presenting for the first time to their GP. A negative faecal calprotectin value at the start of the study, and a positive endoscopic result at the end of the study, should be considered as a false-negative test result.

APPLICABILITY OF STUDY RESULTS

Problems with the patient selection

Test characteristics should be evaluated in a clinically relevant population.[32] In the DOK study the patients with symptoms suggestive of IBD will be recruited in both primary and secondary/tertiary care. Spectrum bias is to be expected as our patient cohorts will have different characteristics.[32] To reduce the risk of spectrum bias one should ideally only include children who initially presented at the primary care level. The low prior probability in this setting makes such a study design infeasible with considerable financial and logistic problems.[9] We decided to use a pragmatic design, based on the following assumptions: in case of a very low prior probability of IBD, a GP wants to avoid unnecessary referrals. The false-positive rate thus needs to be low. Therefore, we will evaluate specificity of faecal calprotectin in children presenting in primary care. In children with red flags, a GP wants to rule out IBD and minimize false-negative results. Sensitivity will thus be evaluated in children referred to secondary or tertiary care (children with red flags in primary care cohort and referred cohort).

We assume that this sensitivity is a representative estimate for sensitivity measured in children with red flags in primary care. This implies two additional assumptions: 1) in both cohorts the ratio IBD/non-IBD in children with red flags will be comparable (which we will test); 2) children with red flags of both cohorts are comparable (which we will test by comparing the clinical characteristics). In case the children from the referred cohort are more severely ill, sensitivity will be overestimated. Heterogeneity can then be assessed by subgroup analyses of the test performance.

CONCLUSION

Low disease prevalence and lack of uniformity in reference standard in primary care creates methodological challenges in primary care level diagnostic accuracy studies. We presented a pragmatic design in which the magnitude of potential bias will be assessed and controlled. Awareness of the potential biases and its implications allows to discuss possible solutions and to overcome such bias. The validity of diagnostic research at the primary care level may be considerably improved with the proposed design.

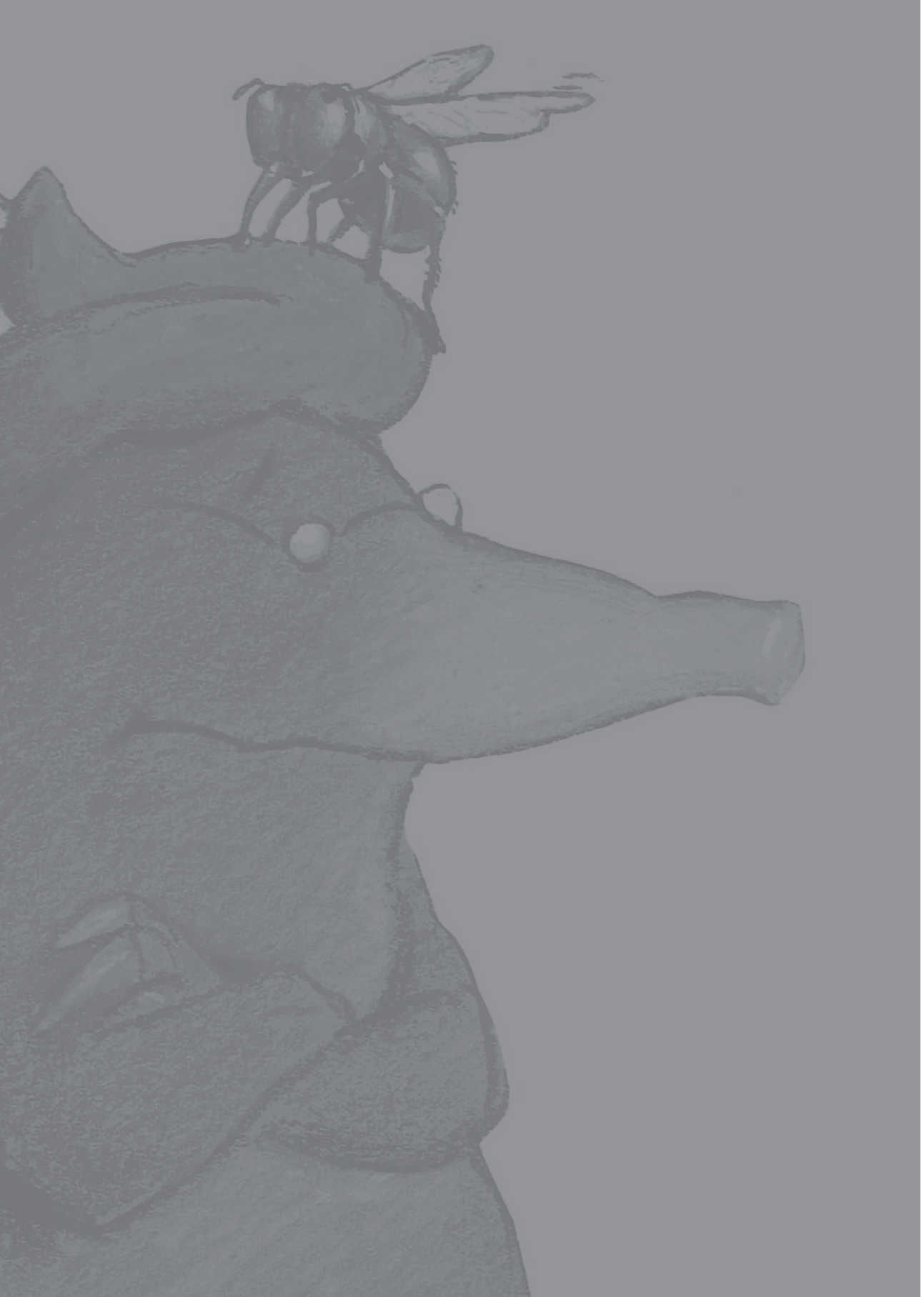
REFERENCES

1. van der Zaag-Loonen HJ, Casparie M, Taminiau JA, Escher JC, Pereira RR, Derkx HH: The incidence of pediatric inflammatory bowel disease in the Netherlands: 1999-2001. *J Pediatr Gastroenterol Nutr* 2004, 38(3):302–307.
2. Gieteling MJ, Lisman-van Leeuwen Y, van der Wouden JC, Schellevis FG, Berger MY: Childhood nonspecific abdominal pain in family practice: incidence, associated factors, and management. *Ann Fam Med* 2011, 9(4):337–343.
3. Rasquin A, Di Lorenzo C, Forbes D, Guiraldes E, Hyams JS, Staiano A, Walker LS: Childhood functional gastrointestinal disorders: child/adolescent. *Gastroenterology* 2006, 130(5):1527–1537.
4. van Rheenen PF, Van de Vijver E, Fidler V: Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: diagnostic meta-analysis. *BMJ* 2010, 341:c3369.
5. Roseth AG, Schmidt PN, Fagerhol MK: Correlation between faecal excretion of indium-111-labelled granulocytes and calprotectin, a granulocyte marker protein, in patients with inflammatory bowel disease. *Scand J Gastroenterol* 1999, 34(1):50–54.
6. Bunn SK, Bisset WM, Main MJ, Gray ES, Olson S, Golden BE: Fecal calprotectin: validation as a noninvasive measure of bowel inflammation in childhood inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 2001, 33(1):14–22.
7. Canani RB, de Horatio LT, Terrin G, Romano MT, Miele E, Staiano A, Rapacciuolo L, Polito G, Bisesti V, Manguso F, Vallone G, Sodano A, Troncone R: Combined use of noninvasive tests is useful in the initial diagnostic approach to a child with suspected inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 2006, 42(1):9–15.
8. Diamanti A, Panetta F, Basso MS, Forgione A, Colistro F, Bracci F, Papadatou B, Francalanci P, Torroni F, Knafelz D, Fina F, Castro M: Diagnostic work-up of inflammatory bowel disease in children: the role of calprotectin assay. *Inflamm Bowel Dis* 2010, 16(11):1926–1930.
9. Buntinx F, Mant D, Van den Bruel A, Donner-Banzhof N, Dinant GJ: Dealing with low-incidence serious diseases in general practice. *Br J Gen Pract* 2011, 61(582):43–46.
10. Levine A, Griffiths A, Markowitz J, Wilson DC, Turner D, Russell RK, Fell J, Ruemmele FM, Walters T, Sherlock M, Dubinsky M, Hyams JS: Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflamm Bowel Dis* 2011, 17(6):1314–1321.
11. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM: QUADAS-2 Group. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Intern Med* 2011, 155(8):529–536.
12. Knottnerus JA, Dinant GJ: Medicine based evidence, a prerequisite for evidence based medicine. *BMJ* 1997, 315(7116):1109–1110.
13. Jellema P, van Tulder MW, van der Horst HE, Florie J, Mulder CJ, van der Windt DA: Inflammatory bowel disease: a systematic review on the value of diagnostic testing in primary care. *Colorectal Dis* 2011, 13(3):239–254.
14. Lewis SJ, Heaton KW: Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol* 1997, 32(9):920–924.
15. Dutch association for pediatrics: *CBO Dutch evidence based guidelines on the diagnosis and treatment of IBD in children. Utrecht: Van Zuiden Communications*; 2008.
16. Baber KF, Anderson J, Puzanovova M, Walker LS: Rome II versus Rome III classification of functional

4

gastrointestinal disorders in pediatric chronic abdominal pain. *J Pediatr Gastroenterol Nutr* 2008, 47(3):299–302.

17. Caplan A, Walker L, Rasquin A: Development and preliminary validation of the questionnaire on pediatric gastrointestinal symptoms to assess functional gastrointestinal disorders in children and adolescents. *J Pediatr Gastroenterol Nutr* 2005, 41(3):296–304.
18. Caplan A, Walker L, Rasquin A: Validation of the pediatric Rome II criteria for functional gastrointestinal disorders using the questionnaire on pediatric gastrointestinal symptoms. *J Pediatr Gastroenterol Nutr* 2005, 41(3):305–316.
19. de Boer RF, Ott A, Kesztyus B, Kooistra-Smid AM: Improved detection of five major gastrointestinal pathogens by use of a molecular screening approach. *J Clin Microbiol* 2010, 48(11):4140–4146.
20. Roseth AG, Fagerhol MK, Aadland E, Schjonsby H: Assessment of the neutrophil dominating protein calprotectin in feces. A methodologic study. *Scand J Gastroenterol* 1992, 27(9):793–798.
21. Ton H, Brandsnes, Dale S, Holtlund J, Skuibina E, Schjonsby H, Johne B: Improved assay for fecal calprotectin. *Clin Chim Acta* 2000, 29(1-2):41–54.
22. Pels LP, Van de Vijver E, Waalkens HJ, Uitentuis J, Gonera-de Jong G, van Overbeek LA, Norbruis OF, Rings EH, van Rheenen PF: Slow hematological recovery in children with IBD-associated anemia in cases of "expectant management". *J Pediatr Gastroenterol Nutr* 2010, 51(6):708–713.
23. Van de Vijver E, Schreuder AB, Cnossen WR, Muller Kobold AC, van Rheenen PF, North Netherlands Pediatric IBD Consortium: Safely ruling out inflammatory bowel disease in children and teenagers without referral for endoscopy. *Arch Dis Child* 2012, 97(12):1014–1018.
24. Mack DR, Langton C, Markowitz J, LeLeiko N, Griffiths A, Bousvaros A, Evans J, Kugathasan S, Otley A, Pfefferkorn M, Rosh J, Mezoff A, Moyer S, Oliva-Hemker M, Rothbaum R, Wyllie R, delRosario JF, Keljo D, Lerer T, Hyams J, Pediatric Inflammatory Bowel Disease Collaborative Research Group: Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics* 2007, 119(6):1113–1119.
25. Bremner A, Roked S, Robinson R, Phillips I, Beattie M: Faecal calprotectin in children with chronic gastrointestinal symptoms. *Acta Paediatr* 2005, 94(12):1855–1858.
26. Carroccio A, Iacono G, Cottone M, Di Prima L, Cartabellotta F, Cavataio F, Scalici C, Montalto G, Di Fede G, Rini G, Notarbartolo A, Averna MR: Diagnostic accuracy of fecal calprotectin assay in distinguishing organic causes of chronic diarrhea from irritable bowel syndrome: a prospective study in adults and children. *Clin Chem* 2003, 49(6 Pt 1):861–867.
27. Fagerberg UL, Loof L, Myrdal U, Hansson LO, Finkel Y: Colorectal inflammation is well predicted by fecal calprotectin in children with gastrointestinal symptoms. *J Pediatr Gastroenterol Nutr* 2005, 40(4):450–455.
28. Buderer NM: Statistical methodology: I Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996, 3(9):895–900.
29. de Groot JA, Dendukuri N, Janssen KJ, Reitsma JB, Bossuyt PM, Moons KG: Adjusting for differential-verification bias in diagnostic-accuracy studies: a Bayesian approach. *Epidemiology* 2011, 22(2):234–241.
30. Naaktgeboren CA, de Groot JAH, van Smeden M, Moons KGM, Reitsma JB: Evaluating Diagnostic Accuracy in the Face of Multiple Reference Standards. *Ann Intern Med* 2013, 159(3):195–202.
31. Reitsma JB, Rutjes AW, Khan KS, Coomarasamy A, Bossuyt PM: A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol* 2009, 62(8):797–806.
32. Mulherin SA, Miller WC: Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002, 137(7):598–602.



CHAPTER 5

DIAGNOSTIC ACCURACY OF FAECAL CALPROTECTIN FOR PAEDIATRIC INFLAMMATORY BOWEL DISEASE IN PRIMARY CARE, A PROSPECTIVE COHORT STUDY

Gea A Holtman

Yvonne Lisman-van Leeuwen

Boudewijn J Kollen

Obbe F Norbruis

Johanna C Escher

Angelika Kindermann

Yolanda B de Rijke

Patrick F van Rheenen

Marjolein Y Berger

Annals of Family Medicine; in press

ABSTRACT

PURPOSE

In specialist care, faecal calprotectin (FCal) is a commonly used non-invasive diagnostic test to rule out inflammatory bowel disease (IBD) in children with chronic gastrointestinal symptoms. The aim of this study was to evaluate the diagnostic accuracy of FCal for IBD in symptomatic children in primary care.

METHODS

We studied 2 prospective cohorts of children with chronic diarrhoea, recurrent abdominal pain, or both: children initially seen in primary care (primary care cohort) and children referred to specialist care (referred cohort). FCal (index test) was measured at baseline and compared with 1 of 2 reference standards for IBD: endoscopic assessment or 1-year follow-up. Physicians were blinded to FCal results, and values greater than 50 µg/g faeces were considered positive. We determined specificity in primary care cohort and sensitivity in referred cohort.

RESULTS

None of the 114 children included in the primary care cohort ultimately received diagnosis of IBD. The specificity of FCal in the primary care cohort was 0.87 (95% CI, 0.80-0.92). Among the 90 children in the referred cohort, 17 patients (19%) ultimately received a diagnosis of IBD. The sensitivity of FCal in the referred cohort was 0.99 (95% CI, 0.81-1.00).

CONCLUSIONS

The findings of this study suggest that a positive FCal result in children with chronic gastrointestinal symptoms seen in primary care is not likely to be indicative of IBD. However, a negative FCal result is likely to be a true negative, which safely rules out IBD in children in whom a general practitioner considers referral to specialist care.

INTRODUCTION

General practitioners (GPs) frequently manage recurrent abdominal pain or diarrhoea in children. These symptoms account for approximately 2% to 5% of all childhood consultations.[1-3] Although they are typically functional in origin, it is essential that organic disease be ruled out. An organic disease that GPs should not miss is inflammatory bowel disease (IBD), that is, Crohn disease and ulcerative colitis. Delay in diagnosing IBD, and the resultant delay in receipt of appropriate treatment, may prolong suffering and can lead to complications such as anaemia, irreversible growth failure, and delayed sexual maturation.[4,5]

According to guidelines, GPs should refer children with chronic diarrhoea, recurrent abdominal pain, or both for specialist care if red flags are present.[6,7] However, the red flags are nonspecific and discriminate poorly between functional and organic gastrointestinal diseases,[8-10] often leading to referral and extensive diagnostic testing. For children with functional disorders, referral or extensive testing may delay appropriate interventions and further decrease well-being.[11,12]

Calprotectin is a calcium-binding protein released from neutrophils during intestinal inflammation that can be easily measured in feaces.[13,14] In specialist care, evidence shows it to be a useful, simple, non-invasive test that can rule out IBD in children with gastrointestinal symptoms.[15-17] However, the diagnostic accuracy of faecal calprotectin (FCal) has not been assessed in children evaluated in primary care.[10,18,19] Primary and specialist care often have different populations, case mixes, and disease severity, which can affect the pre-test probability of IBD and the sensitivity and specificity of the FCal test. The diagnostic accuracy of FCal in the primary care setting should therefore be clarified before this test is recommended for routine use in primary care. We set out to study the diagnostic accuracy of FCal for identifying IBD in children with chronic gastrointestinal symptoms in primary care.

METHODS

STUDY DESIGN

This was a prospective cohort study with a delayed-type cross-sectional design.[20] Children in the Netherlands with chronic gastrointestinal symptoms were included from July 2011 to July 2013 and had 12 months of follow-up. We studied 2 cohorts: 1) the primary care cohort consisted of consecutive children who were seen by any of 64 GPs (38 practices); 2) the referred cohort consisted of consecutive children who were referred for diagnostic work-up by GPs and general paediatricians to any of 4 general hospitals or 3 academic centres, as well as children selected from the primary care cohort based on the presence of at least 1 red flag (Figure 1).

The medical ethics committee of the University Medical Centre Groningen approved the study. Written informed consent was provided by the parents of all children and by all children aged 12 years or older. The study design has been described in more detail elsewhere.[21]

PARTICIPANTS

Children aged 4 to 18 years who sought care for chronic diarrhoea, recurrent abdominal pain,

5

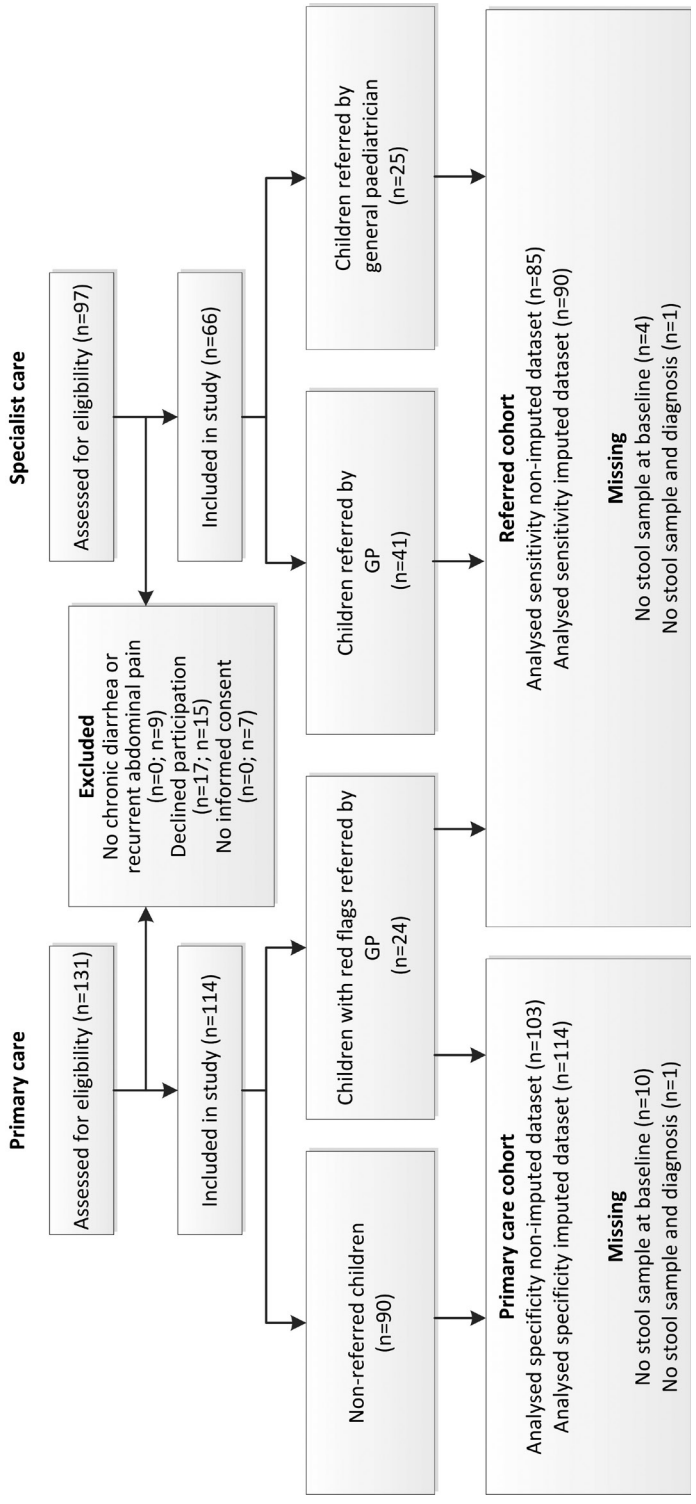


Figure 1. Patient flow in the study.
Note: The group of children who were primarily seen in primary care and selected for referral to specialist care based on ≥1 red flags were evaluated in both analysis.
Faecal calprotectin was not measured in 14 children because no stool sample was collected (9 children), the sample was not stored (2 children), or the sample of faeces was too small to measure the calprotectin value (3 children). Two children with no stool sample were evaluated in both analysis.

Table 1. Definitions for red flags of inflammatory bowel disease.

Red flag	Method of Ascertainment	Definition of positive finding
Alarm symptoms		
Involuntary weight loss	History	Involuntary decrease in weight of >1 kg
Rectal blood loss	History	Rectal blood loss with defecation without constipation according to ROME III criteria
Family history of IBD	History	First-degree relatives
Growth failure	History and physical examination	Target height range > -1 standard deviation score
Extra-intestinal symptoms	Physical examination	Eyes (episcleritis, scleritis, uveitis), skin (erythema nodosum, pyo-derma gangrenosum, psoriasis), mouth ulcers, finger clubbing, arthritis
Peri-anal lesions	Physical examination	skin tags, hemorrhoids, fissures, fistulas, abscess
Blood markers		
Haemoglobin	Local laboratory	4-12 years <7.1 mmol/L, boys 12-18 years <8.1 mmol/L, girls 12-18 years <7.4 mmol/L[39]
C-reactive protein	Local laboratory	>10 mg/L[40]
Erythrocyte sedimentation rate	Local laboratory	>20 mm/h[40]
Platelet count	Local laboratory	>450 $\times 10^9$/L[41]

Abbreviations: IBD, inflammatory bowel disease.

or both were eligible. Chronic diarrhoea was defined as soft to watery stool (score of 5, 6, or 7 on the Bristol stool chart[22]) for at least 2 weeks or at least 2 episodes in the past 6 months. Recurrent abdominal pain was defined as at least 2 episodes of abdominal pain or discomfort in the past 6 months. Children were excluded if they had a previous diagnosis of chronic organic gastrointestinal disease; an evaluation with endoscopy or FCal for gastrointestinal symptoms in 6 months before this study; or difficulty in understanding questionnaires. Furthermore, we excluded children with long-term use (>3 months) of antibiotics, non-steroid anti-inflammatory drugs, or oral corticosteroids in the past 6 months, as well as those aged younger than 4 years, because previous studies have demonstrated elevated calprotectin concentrations in these groups.[23,24]

BASELINE EVALUATION

All participating physicians assessed children for the presence of 10 red flags suggestive of IBD using a structured evaluation form. These red flags consisted of 6 alarm symptoms and 4 blood markers (Table 1). Faeces were tested for pathogens -*Salmonella enterica*, *Campylobacter*

jejuni, Shigella spp/enteroinvasive Escherichia coli (EIEC), and *Shiga toxin-producing Escherichia coli* (STEC)- and for parasites -*Giardia lamblia, Cryptosporidium spp, Dientamoeba fragilis*, and *Entamoeba histolytica*- using real-time polymerase chain reaction.

PATIENT FLOW

All children in the primary care cohort were evaluated by their GP. Children were selected for further diagnostic work-up based on the presence of at least 1 red flag. Children in the referred cohort were evaluated by a paediatric gastroenterologist, who decided whether the child required endoscopic evaluation. All children were followed up for 1 year, during which time the attending physician was free to refer a child for further diagnostic work-up. At 1 year, the physician did a structured physical examination of all children with persisting gastrointestinal symptoms who had not received a diagnosis of IBD. Children with at least 1 red flag at this time were seen by a paediatric gastroenterologist for further evaluation. The attending physicians of children lost to follow-up were contacted after 1 year to provide updated information on persisting symptoms and additional diagnoses.

CONFIRMATORY DIAGNOSIS

IBD was diagnosed by esophagogastroduodenoscopy, ileocolonoscopy, and histopathological examination according to the Porto criteria.[6] A negative endoscopy was defined as the absence of endoscopic and histopathological findings of IBD. Children were considered not to have IBD if, after 12 months, the attending physician found no red flags or the paediatric gastroenterologist decided that red flags were not related to IBD.

INDEX TEST

We used as the index test an FCal test with a lowest sensitivity of 19.5 µg/g faeces and a cut-off point of greater than 50 µg/g faeces according to the manufacturer. Stool samples collected at baseline were stored at -80°C and analysed by a commercially available quantitative enzyme-linked immunosorbent assay (Phi-Cal test, Calpro AS) at the end of the data collection period (September 2014) in the department of clinical chemistry at Erasmus MC. All physicians, pathologists, researchers, and patients were blinded to the outcome of the FCal test and were not allowed to apply for another FCal test during the data collection period. The laboratory staff evaluating FCal were blinded to the children's clinical characteristics and diagnoses.

STATISTICAL ANALYSIS

Very few children seen in primary care with chronic gastrointestinal symptoms have IBD. When introducing new tests in these children, it is important that false-positive outcomes are minimized to avoid referrals for endoscopies. We therefore estimated specificity with adequate precision of FCal measurements in the primary care cohort. Given an expected specificity of 93%,[18,25-27] a 95% confidence interval (CI) and precision of 5%, an IBD prevalence of 5%, and a loss to follow-up of 10%, we needed to include 118 children in the primary care cohort.

The likelihood of IBD increases in children in whom the GP considers a referral. In this group, it is important not to miss a case of IBD, and a low false-negative rate is preferred. We therefore estimated sensitivity with adequate precision of FCal measurement in the referred cohort.

Table 2. Baseline characteristics of primary care cohort and referred cohort.

Characteristics	Main analysis		Referred cohort by origin		Referred cohort by referral	
	Primary care cohort (n = 114)	Referred cohort (n = 90)	Primary care patients with red flag(s) (n = 24)[a]	Specialist care patients (n = 66)	Referred by GP (n = 65)[b]	Referred by general paediatrician (n = 25)
Male (n (%))	38 (33)	37 (41)	8 (33)	29 (44)	29 (45)	8 (32)
Age at baseline (median, IQR)	9 (6-12)	11 (7-15)	9 (6-14)	12 (7-15)	10 (7-14)	14 (10-15.5)
Presenting symptoms (n (%))						
- Recurrent abdominal pain	88 (77)	58 (64)	16 (67)	42 (64)	38 (59)	20 (80)
- Chronic diarrhoea	74 (65)	62 (69)	17 (71)	45 (68)	40 (62)	22 (88)
- Involuntary weight loss	5 (4)[c]	23 (26)	1 (4)	22 (33)	10 (15)	13 (52)
- Rectal blood loss	7 (6)	27 (30)	6 (25)	21 (32)	13 (20)	14 (56)
- Family history of IBD	5 (4)	11 (12)	5 (21)	6 (9)	9/64 (14)	2 (8)
- Growth failure	4 (3)	6 (7)	3 (13)	3 (5)	6 (9)	0 (0)
- Extra-intestinal symptoms	0 (0)	13 (14)	0 (0)	13 (20)	4 (6)	9 (36)
- Peri-anal lesions	7 (6)	13 (14)	7 (29)	6 (9)	9 (14)	4/24 (17)
Positive Blood markers (n/N (%))						
- Haemoglobin[d]	1/111 (1)	11/86 (13)	1/24 (4)	10/62 (16)	5/61 (8)	6/25 (24)
- CRP (>10 mg/L)	2/110 (2)	10/76 (13)	2/24 (8)	8/52 (15)	5/56 (9)	5/20 (25)
- ESR (>20 mm/h)	5/111 (5)	16/83 (19)	5/24 (21)	11/59 (19)	8/59 (14)	8/24 (33)
- Platelet count (>450 $\times 10^9$/L)	4/111 (4)	7/86 (8)	2/24 (8)	5/62 (8)	4/61 (7)	3/25 (12)
Anti-tissue transglutaminase[e] (n/N)	0/100	0/72	0/21	0/51	0/55	0/18
≥1 Red flags[f] (n (%))	29 (25)	68 (76)	24 (100)	44 (67)	43 (66)	25 (100)
Endoscopy (n (%))	2 (2)	29 (32)	2 (8)	27 (41)	9 (14)	20 (80)
IBD (n (%))	0	17 (19)	0	17 (26)	5/64 (8)	12 (48)

Abbreviations: CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; GP, general practitioner; IBD, inflammatory bowel disease.

[a] Five children with red flags were finally not seen by paediatric gastroenterologist: 3 declined because of reduced symptoms, 2 were lost to follow-up.

[b] Including GPs who did not participate in this study.

[c] Three children had no further weight loss after 3 weeks.

[d] Age and sex specific: aged 4-12 years <7.1 mmol/L; aged 12-18 years: boys <8.1 mmol/L, girls <7.4 mmol/L.

[e] Twenty-five children had IgA deficiency.

[f] Red flags: growth failure, involuntary weight loss, rectal blood loss, family history of IBD, extra-intestinal symptoms, peri-anal lesions, positive blood markers (Hb, CRP, ESR, platelet count).

Given an expected sensitivity of 95%,[18,25,27] a 95% CI and precision of ±10%, an IBD prevalence of 20%, and a loss to follow-up of 10%, we needed to include 100 children in referred cohort.[21]

For secondary outcomes, we calculated specificity, post-test probability, and area under the receiver operating characteristic curve (AUC) with 95% CI in the referred cohort. We also determined the effect of different FCal cut-off values (>50 µg/g, >100 µg/g, >250 µg/g)[16,28] on the test characteristics, number of referrals, and missed diagnoses of IBD.

We assumed that children included in specialist care were comparable to those with at least 1 red flag from the primary care cohort. To test this assumption, we compared the characteristics of these groups. In order to evaluate spectrum bias (whereby the test setting affects the test performance), we performed subgroup analyses in the referred cohort,

Table 3. Prevalence of symptoms, blood marker positivity, and FCal positivity by final diagnosis.

Diagnosis	N (%)	symptom positive[a]	Blood marker positive[b]	FCal >50 µg/g	Range FCal (µg/g)
Primary care cohort					
Functional gastrointestinal disorder	108 (95)	24	9/104	12/98	20-257
Gastroenteritis[c]	5 (45)	0	0	1	20-88
Refused endoscopy	1 (1)	1	0	-	-
Referred cohort					
IBD					
Crohn's disease	7 (8)	7	7	6/6	152-2823
Ulcerative colitis	8 (9)	7	4/7	8	53-916
IBD unclassified	2 (2)	2	1	2	79-778
Non-IBD					
Functional gastrointestinal disorder	66 (73)	40	12/87	10/63	20-185
Gastroenteritis[c]	3 (3)	1	0	0	20-45
Reflux esophagitis	1 (1)	0	0	0	22
Celiac disease	1 (1)	1	0	0	20
Solitary rectum ulcer	1 (1)	1	0	1	299

Abbreviations: FCal, faecal calprotectin; IBD, inflammatory bowel disease.

[a] Presence of 1 or more of the following: growth failure, involuntary weight loss, rectal blood loss, extra-intestinal symptoms, peri-anal lesions, family history of IBD.

[b] Blood markers: Haemoglobin (4-12 years <7.1 mmol/L; 12-18 years: boys <8.1 mmol/L, girls <7.4 mmol/L), C-reactive protein (>10 mg/L), erythrocyte sedimentation rate (>20 mm/h), platelet count (>450 x10^9/L).

[c] Gastroenteritis due to *salmonella enteric* (0 cases included by GP; 2 cases included by paediatrician), *Shiga Toxicogene Escherichia Coli* (STEC) (1 and 0), *Giardia lamblia* (4 and 1)

Note: One child refused endoscopy and evaluation of red flags at 12 months' follow-up, so the diagnosis was unknown. Nine children without IBD, including 1 child with a solitary rectal ulcer, underwent upper and lower endoscopy, including ileal intubation. The remaining 3 children did not undergo complete endoscopic evaluation for various reasons: the colonoscopy was prematurely terminated because of mucosal bleeding in 1 child with a functional gastrointestinal disorder, but was not repeated because their symptoms subsided; 1 child with a functional gastrointestinal disorder underwent colonoscopy only, but not esophagogastroduodenoscopy; and 1 child received a diagnosis of celiac disease by esophagogastroduodenoscopy only.

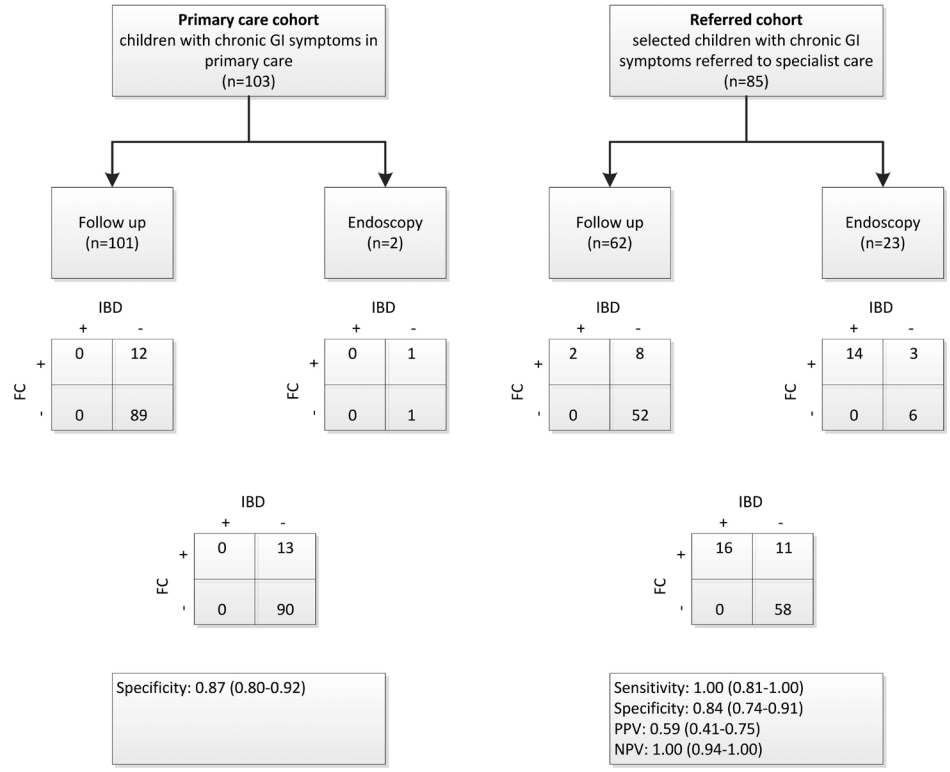


Figure 2. Flow charts and contingency tables for the calculation of diagnostic accuracy in the primary care cohort and referred cohort, using the non-imputed dataset.

Abbreviations: FC, Faecal calprotectin; GI, gastrointestinal; IBD, inflammatory bowel disease; PPV, positive predictive value; NPV, negative predictive value

Note: The left flow chart shows the specificity of FCal (>50 µg/g) for IBD in primary care cohort (11 missing values). Specificity of standard follow-up and endoscopy was 0.88 (95% CI, 0.80–0.93) and 0.50 (95% CI, 0.09–0.91), respectively. The right flow chart shows the test characteristics of FCal (>50 µg/g) for IBD in referred cohort (5 missing values). Sensitivity of reference standard follow-up and endoscopy were 1.00 (95% CI, 0.34–1.00) and 1.00 (95% CI, 0.78–1.00), respectively; values of specificity were 0.87 (95% CI, 0.76–0.93) and 0.67 (95% CI, 0.35–0.88), respectively.

comparing children who were referred by a GP with those referred by a general paediatrician. To evaluate the likelihood of differential verification bias,[29] we separately calculated the test characteristics for both endoscopy and 12-month follow-up.

We conducted a missing value analysis to rule out "missing not at random" as possible explanation for missing data from the variables FCal and diagnosis. The missing data were replaced based on a multiple imputation procedure (conditional specification, predictive mean matching,[30] 20 iterations, and 20 datasets). The patient characteristics, all symptoms, all diagnostic tests, setting, endoscopy performed, and whether IBD was diagnosed were

Table 4. Test characteristics at increasing calprotectin cut-off levels in the referred cohort using the imputed dataset (n=90).

Test characteristics	>50 µg/g	>100 µg/g	>250 µg/g
Sens (95% CI)	0.99 (0.81–1.00)	0.87 (0.65–0.96)	0.81 (0.58–0.93)
Spec (95% CI)	0.84 (0.74–0.91)	0.93 (0.84–0.97)	0.98 (0.92–0.99)
PPV (95% CI)	0.60 (0.42–0.76)	0.74 (0.53–0.88)	0.92 (0.69–0.98)
NPV (95% CI)	1.00 (0.94–1.00)	0.97 (0.89–0.99)	0.96 (0.88–0.98)
Fewer referrals (n (%))[a]	61 (68%)	69 (77%)	74 (82%)
Missed cases of IBD (n (%))[b]	0 (0%)	2 (12%)	3 (18%)

Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

[a] Denominator is the 90 children in the referred cohort

[b] Denominator is the 17 children in the referred cohort ultimately given a diagnosis of IBD

included as predictors. We used Rubin's rule to calculate the pooled AUC.[31] Results of the analyses on the non-imputed and imputed dataset were compared to assess the effect of multiple imputations on diagnostic accuracy. Statistical analyses were performed with IBM SPSS for Windows, Version 20.0 (IBM Corp).

RESULTS

PARTICIPANTS

The primary care cohort had 114 children, and the referred cohort had 90 children (24 with red flags from primary care plus 66 children from specialist care) (Figure 1). Table 2 shows that the 66 children from specialist care more often had weight loss, extra-intestinal symptoms, and decreased haemoglobin levels compared with the 24 children with red flags from primary care. The 25 children who were referred to the paediatric gastroenterologist by a general paediatrician showed a higher IBD prevalence than 65 children who were referred by their GP (12 vs 5 cases).

DIAGNOSES

None of the children in the primary care cohort received a diagnosis of IBD (Table 3). The final diagnosis was based on 12 months' follow-up in 111 children and endoscopy in 2 children. Of the 90 children in the referred cohort, 29 (32%) underwent endoscopic evaluation, and 17 (19%) ultimately received a diagnosis of IBD.

FAECAL CALPROTECTIN

The median intervals from stool collection to endoscopy were 4 days and 8 days for children with IBD and without IBD, respectively; however, 11 of the 27 children (2 had missing samples) who underwent endoscopy experienced a delay of more than 1 month. Two children

in the primary care cohort had similar calprotectin values in samples collected at baseline and shortly before endoscopy. In total, values were missing for 11 (9.6%) children in the primary care cohort and 5 (5.5%) in the referred cohort.

The contingency tables and test characteristics for FCal by follow-up or endoscopy are presented in Figure 2 for the non-imputed dataset. The outcomes were comparable for all parameters in the imputed and non-imputed datasets. In the imputed dataset, the specificity of FCal for IBD in the primary care cohort was 0.87 (95% CI, 0.80–0.92). Six of 24 children with red flags and 9 of 90 children without red flags had a positive calprotectin value. In the referred cohort, the sensitivity was 0.99 (95% CI, 0.81–1.00). The AUC of FCal in referred cohort was 0.98 (95% CI, 0.96–1.00).

In subgroup analyses, the sensitivities of the test in children referred to specialist care by their GP and by a general paediatrician were 0.98 (95% CI, 0.58–1.00) and 1.00 (95% CI, 0.76–1.00), respectively. An increase in the cut-off value for FCal from 50 µg/g to 250 µg/g faeces in the referred cohort would lead to an extra 14% reduction in referrals for diagnostic work-up for IBD, but would also increase the percentage of missed IBD diagnoses from 0% to 18% (Table 4).

DISCUSSION

This is the first study to evaluate the test characteristics of FCal as a marker for IBD in children seen for chronic gastrointestinal symptoms in primary care. None of the children in the primary care cohort ultimately received a diagnosis of IBD, suggesting that children with chronic gastrointestinal symptoms should not be referred directly for evaluation of IBD. Current guidelines recommend referral for diagnostic work-up based on the presence of red flags. In our primary care cohort, referrals based on red flags would have resulted in a higher false-positive referral rate for IBD compared to referrals based on FCal exceeding 50 µg/g faeces. To date, however, there is insufficient evidence to recommend FCal as a tool to guide decisions about referral for diagnostic work-up of all children with chronic gastrointestinal symptoms seen in primary care.

FCal showed high sensitivity (0.99; 95% CI, 0.81–1.00) in the referred cohort. Therefore, a negative FCal may safely rule out IBD and therewith reduce the number of referrals for evaluation of IBD in children whom the GP considers a referral. Nevertheless, the 95% CIs of false-negative rates are large, because of the relatively small numbers of children with IBD included in our study, and should be interpreted with caution. In this study we focused on IBD, but from the point of view of the GP, it is important to determine whether the symptoms are related to any organic disease. In addition to 17 children with IBD, 3 children were found to have other organic diseases (1 had celiac disease, 1 had reflux esophagitis, 1 had solitary rectum ulcer). Children with celiac disease and reflux esophagitis had normal FCal levels and would have been missed if a referral had been solely based on the results of this test.

Cut-off points might need to be higher in primary care to maintain a high negative predictive value.[32] Although we found that an increase of the threshold from 50 µg/g to 250

5

μg/g faeces reduced referrals by 14% (with a drop from 32% to 18% referred), this threshold also led to false-negative results and missed IBDs (with an increase from 0% to 18% cases missed) in the referred cohort. A pragmatic approach may be to monitor children with an initial calprotectin value between 50 μg/g and 250 μg/g faeces. Children whose symptoms persist and whose calprotectin values remain high can still be referred later. A similar approach has been suggested for adults in primary care, where it was suggested that patients with irritable bowel syndrome and an initial FCal value between 50 μg/g and 150 μg/g faeces who had persistent symptoms without treatment should be re-tested after 3 months.[28]

The specificity of 0.87 that we identified in the primary care cohort is higher than that reported in studies performed in specialist care, where the pooled specificity ranged between 0.68 and 0.76.[15-17] We expected lower specificities because the patient mix was thought to be more diverse in primary care and because calprotectin concentrations increase in conditions such as gastroenteritis.[33] Moreover, in the referred cohort, the specificity of FCal was lower in children who underwent endoscopy (0.67) than in those who received clinical follow-up (0.87). The higher specificity might be explained by higher numbers of children with functional disease in the primary care cohort and in children with clinical follow-up in the referred cohort. Consequently, the test setting (primary versus specialist care) might affect the specificity of FCal (spectrum bias).

We found a sensitivity of 0.99 in the referred cohort, which is comparable to that reported in other studies performed in specialist care (IBD prevalence of about 60%), where the pooled sensitivity ranged between 0.92 and 0.98.[15-17] The reported sensitivity of FCal in this study might be an overestimation of that in children in whom a GP considers a referral for diagnostic work-up. However, in the subgroup of children referred by a GP (IBD prevalence of 8%) the sensitivity was comparable to that of children referred by a general paediatrician (IBD prevalence of 48%) (0.98 vs 1.00). We therefore assume that spectrum bias did not substantially affect the sensitivity of FCal. Lack of spectrum bias might be explained by the fact that the intestinal inflammation caused by IBD is usually severe enough to increase the calprotectin value to more than 50 μg/g faeces, even in the early stages of the disease. A meta-analysis showed that sensitivity remains stable over a range of prevalences and was not substantially influenced by spectrum bias.[34] A normal FCal level thus could be used to prevent a referral of children with functional symptoms to specialist care.

Ideally, a study of FCal would include consecutive children with symptoms suggestive of IBD initially evaluated in primary care. However, this design would be extremely time consuming and costly. In order to reflect current daily, real-world practice, we included a cohort of children in whom the GP considered a referral for diagnostic work-up for chronic gastrointestinal symptoms. We assumed that the children first seen in specialist care were comparable to the selected children with red flags seen in the primary care cohort. We are confident that this assumption is valid as there were only a few differences in characteristics between these groups. However, children referred by a GP had a lower probability of IBD (8%) than those referred by a general paediatrician (48%). These findings are consistent with what one might expect in the Dutch healthcare system, where children can consult a paediatrician only after obtaining a referral from their GP, and a paediatric gastroenterologist can be consulted after a referral from a GP or general paediatrician. Comparable healthcare systems exist in the United Kingdom, Scandinavia, Canada, New Zealand, and Australia.[35]

Not all patients received the same reference standard test, which may cause differential verification bias.[29] As it is unethical to perform endoscopy in children with a low likelihood of organic gastrointestinal disease, these children received follow-up evaluations during 1 year. An important aspect for deciding whether the verification lead to biased estimates of accuracy is the length of the follow-up period. We are confident that we identified all children with IBD, because of the extremely low probability that a child without red flags or indication for endoscopy during 1 year of follow-up has the disease.[36] We did not identify new cases of IBD at the end of the follow-up period, even among children who developed red flags. Only 2 children ultimately received a diagnosis of IBD within the follow-up period (both before 6 months). The use of follow-up in children in whom endoscopy is not considered ethical does not correspond very well with the ideal situation that arise when the diagnosis of all children is determined by endoscopy. However, following children during 1 year is the best option given the reality of clinical care.[20]

To recommend a test in a new setting, the diagnostic value of that test needs to be investigated in that setting.[37] We found that, in selected children in whom a GP considers a referral, FCal has satisfactory discriminating power between children with and without IBD. However, of greater clinical relevance, is whether FCal can add to the diagnostic information that is readily available from a thorough history and physical examination.[38] Moreover, the added value of commonly used blood markers should be compared with the added value of FCal. Further research is therefore needed to determine whether FCal should be incorporated into the routine diagnostic evaluation of paediatric patients with chronic gastrointestinal symptoms and red flags in primary care. In addition, research should be performed to evaluate the cost-effectiveness of FCal in primary care.

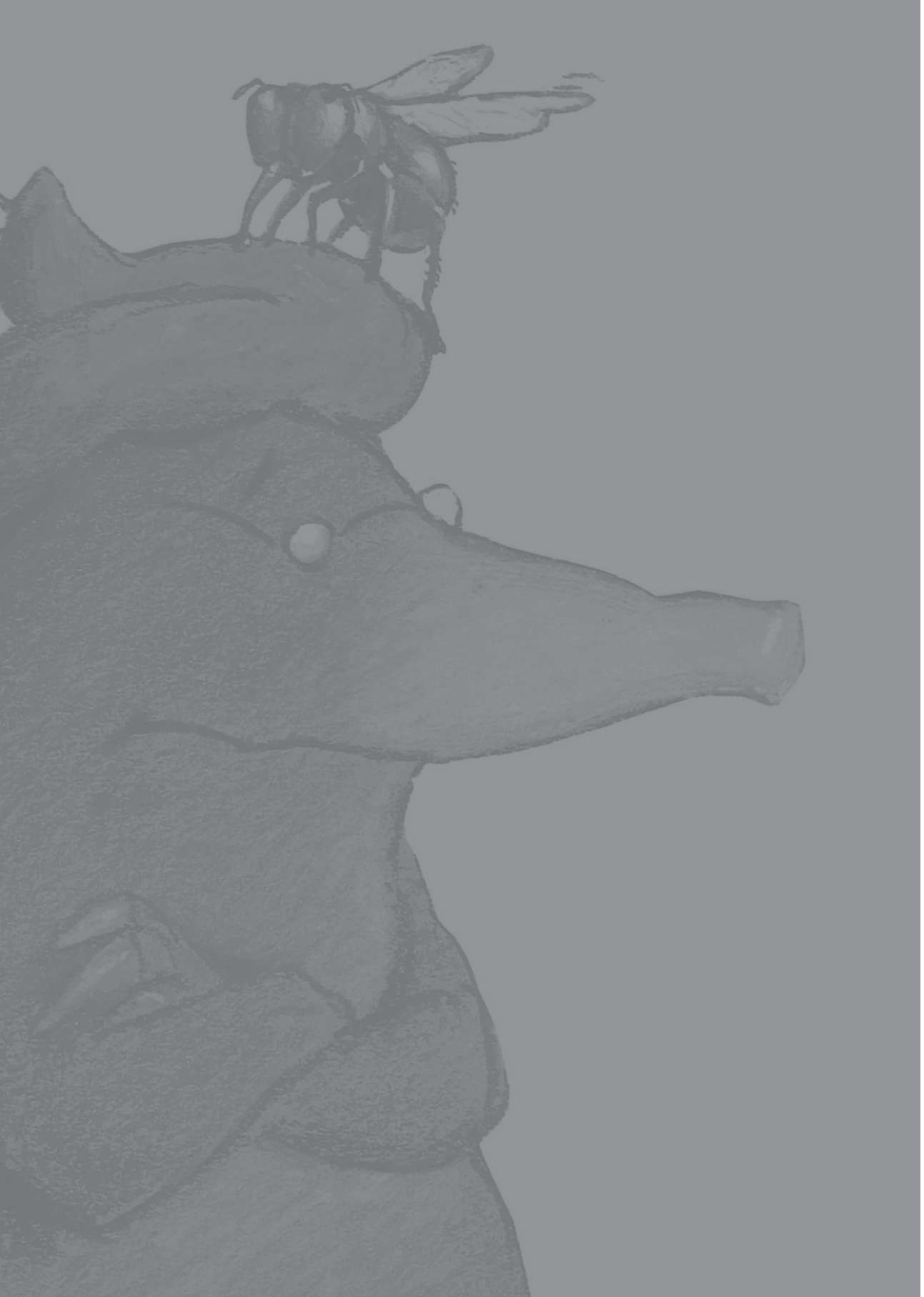
5

REFERENCES

1. Gieteling MJ, Lisman-van Leeuwen Y, van der Wouden JC, Schellevis FG, Berger MY. Childhood nonspecific abdominal pain in family practice: Incidence, associated factors, and management. *Ann Fam Med*. 2011;9(4):337-343.
2. BEACH Program, AIHW General Practice Statistics and Classification Unit. Presentations of abdominal pain in australian general practice. *Aust Fam Physician*. 2004;33(12):968-969.
3. Starfield B, Hoekelman RA, McCormick M, et al. Who provides health care to children and adolescents in the united states? *Pediatrics*. 1984;74(6):991-997.
4. Spray C, Debelle G, Murphy M. Current diagnosis, management and morbidity in paediatric inflammatory bowel disease. *Acta Paediatrica*. 2001;90(4):400-405.
5. Kim SC, Ferry GD. Inflammatory bowel diseases in pediatric and adolescent patients: Clinical, therapeutic, and psychosocial considerations. *Gastroenterology*. 2004;126(6):1550-1560.
6. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr*. 2014;58(6):795-806.
7. de Ridder L, Rings EH, Escher JC, CBO-werkgroep 'IBD bij kinderen'. Guideline 'diagnosis and treatment of inflammatory bowel disease in children'. *Ned Tijdschr Geneeskd*. 2010;154:A1898.
8. Cabrera-Abreu JC, Davies P, Matek Z, Murphy MS. Performance of blood tests in diagnosis of inflammatory bowel disease in a specialist clinic. *Arch Dis Child*. 2004;89(1):69-71.
9. Quail MA, Russell RK, Van Limbergen JE, et al. Fecal calprotectin complements routine laboratory investigations in diagnosing childhood inflammatory bowel disease. *Inflamm Bowel Dis*. 2009;15(5):756-759.
10. Henderson P, Casey A, Lawrence SJ, et al. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease. *Am J Gastroenterol*. 2012;107(6):941-949.
11. Gieteling MJ, Bierma-Zeinstra SM, Passchier J, Berger MY. Prognosis of chronic or recurrent abdominal pain in children. *J Pediatr Gastroenterol Nutr*. 2008;47(3):316-326.
12. van Tilburg MA, Venepalli N, Ulshen M, Freeman KL, Levy R, Whitehead WE. Parents' worries about recurrent abdominal pain in children. *Gastroenterology Nursing*. 2006;29(1):50-55.
13. Fagerhol MK, Dale I, Anderson T. Release and quantitation of a leucocyte derived protein (L1). *Scand J Haematol*. 1980;24(5):393-398.
14. Røseth A, Fagerhol M, Aadland E, Schjønsby H. Assessment of the neutrophil dominating protein calprotectin in feces: A methodologic study. *Scand J Gastroenterol*. 1992;27(9):793-798.
15. van Rheenen PF, Van de Vijver E, Fidler V. Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: Diagnostic meta-analysis. *BMJ*. 2010;341:c3369.
16. Henderson P, Anderson NH, Wilson DC. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease: A systematic review and meta-analysis. *Am J Gastroenterol*. 2014;109(5):637-645.
17. Degraeuwe PL, Beld MP, Ashorn M, et al. Faecal calprotectin in suspected paediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2015;60(3):339-346.
18. Fagerberg UL, Loof L, Myrdal U, Hansson LO, Finkel Y. Colorectal inflammation is well predicted by fecal calprotectin in children with gastrointestinal symptoms. *J Pediatr Gastroenterol Nutr*. 2005;40(4):450-455.
19. Holtman GA, Lisman-van Leeuwen Y, Reitsma JB, Berger MY. Noninvasive tests for inflammatory bowel disease: A meta-analysis. *Pediatrics*. 2016;137(1):1-11.
20. Knottnerus JA, Dinant GJ. Medicine based evidence, a prerequisite for evidence based medicine. *BMJ*. 1997;315(7116):1109-1110.
21. Holtman GA, Lisman-van Leeuwen Y, Kollen BJ, et al. Challenges in diagnostic accuracy studies in primary care: The fecal calprotectin example. *BMC Fam Pract*. 2013;14:179-2296-14-179.
22. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol*. 1997;32(9):920-924.
23. Oord T, Hornung N. Fecal calprotectin in healthy children. *Scand J Clin Lab Invest*. 2014;74(3):254-258.
24. Gisbert J, McNicholl A. Questions and answers on the role of faecal calprotectin as a biological marker in inflammatory bowel disease. *Digestive and Liver Disease*. 2009;41(1):56-66.
25. Canani RB, de Horatio LT, Terrin G, et al. Combined use of noninvasive tests is useful in the initial diagnostic approach to a child with suspected inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2006;42(1):9-15.
26. Bremner A, Roked S, Robinson R, Phillips I, Beattie M. Faecal calprotectin in children with chronic gastrointestinal symptoms. *Acta Paediatr*. 2005;94(12):1855-1858.
27. Carroccio A, Iacono G, Cottone M, et al. Diagnostic accuracy of fecal calprotectin assay in distinguishing organic causes of chronic diarrhea from irritable bowel syndrome: A prospective study in adults and children. *Clin Chem*. 2003;49(6 Pt 1):861-867.
28. Pavlidis P, Chedgy FJ, Tibble JA. Diagnostic accuracy and clinical application of faecal calprotectin in adult patients presenting with gastrointestinal symptoms in primary care. *Scand J Gastroenterol*. 2013;48(9):1048-1054.
29. Naaktgeboren CA, De Groot JA, van Smeden M, Moons KG, Reitsma JB. Evaluating diagnostic accuracy in the face of multiple reference standards. *Ann Intern Med*. 2013;159(3):195-202.
30. Vink G, Frank LE, Pannekoek J, Buuren S. Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*. 2014;68(1):61-90.
31. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines. *BMC Med Res Methodol*. 2009;9:57-2288-9-57.
32. Logan R. Faecal calprotectin for the diagnosis of inflammatory bowel disease. *BMJ*. 2010;341:c3636.
33. Chen CC, Huang JL, Chang CJ, Kong MS. Fecal calprotectin as a correlative marker in clinical severity of infectious diarrhea and usefulness in evaluating bacterial or viral pathogens in children. *J Pediatr Gastroenterol Nutr*. 2012;55(5):541-547.
34. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ*. 2013;185(11):E537-44.
35. Thomson S, Osborn R, Squires D, Reed SJ. International profiles of health care systems 2011: Australia, canada, denmark, england, france, germany, iceland, italy, japan, the netherlands, new zealand, norway, sweden, switzerland, and the united states. . 2011.
36. Jellema P, van Tulder MW, van der Horst HE, Florie J, Mulder CJ, van der Windt DA. Inflammatory bowel disease: A systematic review on the value of diagnostic testing in primary care. *Colorectal Dis*. 2011;13(3):239-254.
37. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem*. 2004;50(3):473-476.
38. Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM. Quantifying the added value of a diagnostic test or marker. *Clin Chem*. 2012;58(10):1408-1417.

5

39. Pels LP, Van de Vijver E, Waalkens HJ, et al. Slow hematological recovery in children with IBD-associated anemia in cases of "expectant management". *J Pediatr Gastroenterol Nutr*. 2010;51(6):708-713.
40. Van de Vijver E, Schreuder AB, Cnossen WR, Muller Kobold AC, van Rheenen PF, North Netherlands Pediatric IBD Consortium. Safely ruling out inflammatory bowel disease in children and teenagers without referral for endoscopy. *Arch Dis Child*. 2012;97(12):1014-1018.
41. Mack DR, Langton C, Markowitz J, et al. Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics*. 2007;119(6):1113-1119.



CHAPTER 6

DIAGNOSTIC TEST STRATEGIES FOR INFLAMMATORY BOWEL DISEASE IN CHILDREN PRESENTING AT PRIMARY CARE LEVEL

Gea A Holtman

Yvonne Lisman-van Leeuwen

Boudewijn J Kollen

Obbe F Norbruis

Johanna C Escher

Laurence C Walhout

Angelika Kindermann

Yolanda B de Rijke

Patrick F van Rheenen

Marjolein Y Berger

Submitted

ABSTRACT

BACKGROUND

In children with symptoms suggestive of inflammatory bowel disease (IBD) presenting at primary care level it is unclear what the optimal test strategy is to identify those children who require specialist care. We evaluated three test strategies to determine the optimal referral strategy for specialist care in children suspected of IBD: 1) alarm symptoms, 2) alarm symptoms and C-reactive protein, and 3) alarm symptoms and faecal calprotectin.

METHODS

A prospective cohort study was conducted, including children with chronic gastrointestinal symptoms referred to paediatric gastroenterology (n=65 referred by general practitioners, n=25 referred by general paediatricians). Outcome was defined as IBD confirmed by endoscopy, or IBD ruled out by either endoscopy or unremarkable clinical 12-month follow-up with no indication for endoscopy. Receiver operating characteristics and decision curves were used to compare the three test strategy probabilities generated by logistic regression analyses.

RESULTS

We included 90 children, of whom 17 (19%) had IBD. Adding faecal calprotectin to alarm symptoms increased the area under the curve significantly from 0.80 (0.69–0.90) to 0.97 (0.93–1.00). C-reactive protein, when added to alarm symptoms, did not increase the area under the curve significantly. Decision curves confirmed these patterns and showed that a combination of alarm symptoms and faecal calprotectin is the diagnostic test strategy with the highest net benefit at all reasonable threshold probabilities.

INTERPRETATION

An evaluation of alarm symptoms and faecal calprotectin showed to be the optimal strategy for further stratifying children who have already been identified as at risk for IBD by the general practitioners.

INTRODUCTION

Abdominal pain is the most common gastrointestinal symptom that prompt a clinic visit.[1] The abdominal symptoms are frequently attributed to functional gastrointestinal disorders (FGIDs) and seldom to organic diseases.[2–4] Children with functional gastrointestinal disorders have no structural or biochemical abnormalities and can be managed in primary care.[5,6] In these children, excessive testing will sustain complaints and decreases patients well-being.[7,8] However, a thorough differential diagnosis of organic diseases, like inflammatory bowel disease (IBD), is necessary to avoid delay in diagnosis and appropriate treatment in these children.[9]

Differentiation between functional disorders and IBD, however, is difficult, since symptoms are non-specific and overlapping. The absence of alarm symptoms (e.g. weight loss, rectal blood loss) may assist the general practitioner in ruling out IBD and prevent referrals of children with FGIDs in order to preclude somatic disease.[10] However, these alarm symptoms are nonspecific and common in primary care, and mostly related to the easily recognized less "alarming" illnesses (i.e. rectal bleeding can be caused by fissura ani related to constipation). The trade-off between benefit (early recognition of IBD) and harm (referral of functional gastrointestinal disorders) of a referral strategy and the lack of adequate tools to discriminate children with IBD, triggers general practitioners to either refer children with abdominal complaints for further diagnostic work-up or to perform a variety of non-valid tests.[2]

In the Netherlands, alarm symptoms measured by history and physical examination in combination with blood markers is the commonly used diagnostic strategy to triage those who need a referral for specialist care.[11] C-reactive protein showed the best diagnostic performance of reported blood markers.[12,13] More recently, faecal calprotectin has shown to be a useful, simple, non-invasive test that can exclude IBD.[12,14,15] To date, test characteristics such as sensitivity, specificity and area under the curve are presented. However, the clinical consequences of alarm symptoms, C-reactive protein or faecal calprotectin are unknown.[16]

This study aims to evaluate the added diagnostic performance of C-reactive protein and faecal calprotectin beyond alarm symptoms to determine the optimal diagnostic test strategy for referral for specialist care in children suspected of IBD. For that purpose, we compared three test strategies: 1) alarm symptoms, 2) alarm symptoms and C-reactive protein, and 3) alarm symptoms and faecal calprotectin.

METHODS

SETTING AND PARTICIPANTS

We conducted a prospective cohort study between July 2011 and September 2014, including children recruited consecutively by 38 general practitioner practices, 4 general hospitals, and 3 academic centres in the Netherlands.[17] For inclusion, children had to be aged 4–18 years and presenting with chronic diarrhoea (soft or watery stool, matching scores 5–7 of the Bristol Stool chart, for ≥2 weeks or ≥2 episodes in the past 6 months) 18 or recurrent abdominal pain (≥2 episodes of abdominal pain or discomfort in the past 6 months).[10,11] Children recruited

6

by general practitioners were only included when they were at risk (≥1 alarm symptom or positive blood test) for IBD and referred to a paediatric gastroenterologist at baseline assessment, in accordance with the study protocol.[17] Participants were excluded if they had a known diagnosis of chronic organic gastrointestinal disease; had undergone endoscopic evaluation or faecal calprotectin measurement within the preceding 6 months; or had difficulty in understanding questionnaires. Furthermore, we excluded children with chronic use (>3 months) of antibiotics, non-steroid anti-inflammatory drugs, or oral corticosteroids, as well as children aged under 4 years, because their calprotectin levels have been shown to be higher than those observed in healthy older children and adults.[19,20]

The Medical Ethics Committee of the University Medical Centre Groningen, the Netherlands, approved the study. Parents of all children provided written informed consent, as did children aged 12 years or older.

ALARM SYMPTOMS, BLOOD MARKERS, AND FAECAL CALPROTECTIN

Alarm symptoms and the thresholds for blood markers and faecal calprotectin are presented in Table 1. At baseline the general practitioner or general paediatrician performed a structured physical examination identify the six alarm symptoms using standardized forms. A blood sample was taken from all patients to measure the four blood markers. Directly after baseline, all children collected a stool sample and sent it to a laboratory, where it was stored at −80°C. At the end of the data collection period (September 2014), calprotectin was measured by a commercially available quantitative enzyme-linked immunosorbent assay (Phical test, CALPRO AS, Oslo, Norway). All physicians, researchers, and patients were blinded to the outcome of the faecal calprotectin test, but not to the results of alarm symptoms or blood analyses. The technicians in the laboratory were blinded to the clinical characteristics and diagnoses of patients.

REFERENCE STANDARDS

The diagnosis of IBD was confirmed by esophagogastroduodenoscopy and ileocolonoscopy, with histology, according to the Porto criteria.[10] Absence of IBD was defined as no endoscopic and histopathologic evidence of IBD or no indication for endoscopy within or at 12 months' follow-up. At baseline or during follow-up, the paediatric gastroenterologist performed endoscopy if indicated based on paediatricians global assessment, physical examination, and blood results. After 12 months, the general practitioners or paediatricians of children lost to follow-up were contacted to provide updated information on relevant diagnoses and the child's health status.

STATISTICAL ANALYSES

We calculated the tests characteristics with 95% confidence intervals (CIs). In addition, we calculated the area under the receiver operating characteristic curve (AUC) for blood markers, and faecal calprotectin.

We constructed a basic model predicting the presence of IBD, using logistic regression analysis. The dependent variable was the diagnosis IBD (dichotomous) and for the independent variable we combined alarm symptoms into one variable (continuous) to prevent

Table 1. Definitions for alarm symptoms and diagnostic thresholds for blood markers and faecal calprotectin.

Symptoms or tests	Measurement	Positive
Alarm symptoms		
Involuntary weight loss	History	Involuntary decrease in weight of > 1 kg
Rectal blood loss	History	Rectal bleeding with defecation without constipation according to ROME III criteria
Family history of IBD	History	Affected first-degree relatives
Growth failure	History and physical examination	Target height range > -1 standard deviation score
Extra-intestinal symptoms	Physical examination	Eyes (episcleritis, scleritis, uveitis), skin (erythema nodosum, pyoderma gangrenosum, psoriasis), mouth ulcers, finger clubbing, arthritis
Peri-anal lesions	Physical examination	skin tags, haemorrhoids, fissures, fistulas, abscess
Blood markers		
haemoglobin	Local laboratory	4-12 years < 7.1 mmol/l, boys 12-18 years < 8.1 mmol/l, girls 12-18 years < 7.4 mmol/l[39]
C-reactive protein	Local laboratory	> 10 mg/l[31]
erythrocyte sedimentation rate	Local laboratory	> 20 mm/h[31]
platelet count	Local laboratory	> 450 x10^9/l[40]
Faecal marker		
faecal calprotectin	ELISA (Phical)	> 50 µg/g

Abbreviations: ELISA: enzyme-linked immunosorbent assay; IBD: inflammatory bowel disease

6

an unfavourable number of events per variable.[21] We added weighting scores to the alarm symptoms, because these symptoms have different predictive values.[10,11]

To evaluate the added value of C-reactive protein and faecal calprotectin (both continuous outcome) to the alarm symptoms (basic model), we added these variables to the basic model. The added value of the two models was compared to the basic model by the difference in AUC and was considered significantly different if z was ≥1.96. The paired z score between models was calculated using the equation: $z = AUC_1 - AUC_2 / \sqrt{[SE^2_1 + SE^2_2 - 2rSE_1SE_2]}$, where r was the Spearman correlation between the two models.[22] The models were also assessed with goodness-of-fit tests and calibration plots.

We used decision curve analysis to evaluate the clinical usefulness of decision making based on the three diagnostic strategies. The decision curve analysis weighed the clinical consequence of false positives at different diagnostic probability thresholds.[23–25] The higher the probability threshold, the unnecessary referrals are less acceptable for the general practitioner and the higher is the "weight" assigned to false positives. We choose a range of threshold probabilities with an upper limit of 40%, because it is unrealistic that any general practitioner would need more than a 40% risk of IBD before referral is recommended.[16] The net benefit of making a decision based on a strategy was calculated using the following equation: net benefit = [true positives /n] – [false positives /n]*[Pt / (1–Pt)], where n was the total number of children and Pt was a given threshold probability.[23] Strategies with the highest net benefit at a particular threshold are considered preferable to alternative strategies. The reduction in referrals for further diagnostic work-up per 100 children was calculated with the equation: Δ net benefit / [Pt / (1 – Pt)]*100.[23]

To evaluate the verification pattern of which child received endoscopy at baseline, we compared baseline characteristics and risk of IBD in children with and without endoscopy at baseline. We expected that especially high risk children were subjected to endoscopies.

We described the patterns of missing data by analysing the distribution of variables in those with and without missing data. If the distribution between variables was different, the missing data were missing at random (MAR).[26] We conducted multiple imputations (fully conditional specification, predictive mean matching,[27] 20 iterations, 20 datasets), using the following variables as predictors in the imputation procedure: sex, age, length, weight, chronic abdominal pain, chronic diarrhoea, the six alarm symptoms, all blood markers, endoscopy, setting, faecal calprotectin value, and outcome variable: diagnosis of IBD.[28] We used Rubin's rule to calculate the pooled AUC.[29] Statistical analyses were performed with IBM SPSS version 20.0.0.2 (IBM corp., Armonk, New York, USA) and STATA/SE 13 (Stata Corp, College station, TX, USA).

RESULTS

PARTICIPANTS

Figure 1 summarizes the patient flow in this study. Outcome IBD was based on results of endoscopy in 29 children and on the findings after 12 months' follow-up in 61 children. Children referred to the paediatric gastroenterologist by a general paediatrician (n=25) were

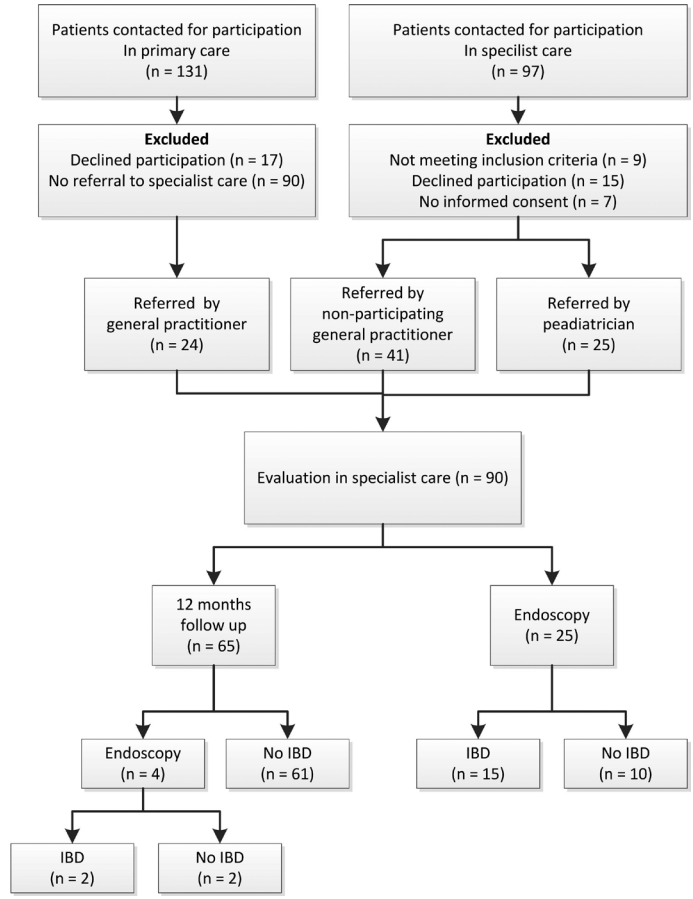


Figure 1. Patient flow.
Abbreviations: IBD: inflammatory bowel disease.

older, had more frequently alarm symptoms, and higher IBD prevalence than children who were referred by a general practitioner (n=65). Children who were subjected to an endoscopy at baseline (n=25) had more frequently alarm symptoms, positive laboratory markers, and high risk for IBD than children who underwent no endoscopy at baseline (n=65) (Table 2).

DIAGNOSES

IBD was confirmed in 17 patients (19%), of whom 7 had Crohn's disease, 8 had ulcerative colitis, and 2 had IBD-unclassified. Of the 72 children (80%) with other diagnoses, 66 (73%) had a functional gastrointestinal disorder, 3 (3%) had gastroenteritis (*salmonella enteric* (n=2); *Giardia lamblia* (n=1)), 1 (1%) had reflux esophagitis, 1 (1%) had celiac disease, and 1 (1%) had a solitary rectal ulcer. In one patient the diagnosis is missing, because the child, aged 16 years, refused both endoscopic evaluation at baseline and evaluation by a general practitioner at 12 months.

Table 2. Baseline characteristics of children referred by their general practitioner or paediatrician and children with and without endoscopy at baseline.

	Referred by general practitioners (n = 65)	Referred by general paediatrician (n = 25)	No endoscopy At baseline (n = 65)	Endoscopy At baseline (n = 25)
Male sex (n (%))	29 (45)	8 (32)	27 (42)	10 (40)
Age in years at baseline (median, IQR)	10 (7-14)	14 (10-15.5)	9 (6-14)	15 (12-16)
Duration symptoms (n (%))				
< 0.5 year	14 (22)	6 (24)	12 (19)	8 (32)
> 1 year	41 (63)	9 (36)	42 (65)	8 (32)
History and physical examination (n (%))				
Growth failure	6 (9.2)	0 (0)	5 (8)	1 (4)
Involuntary weight loss	10 (15)	13 (52)	10 (15)	13 (52)
Rectal blood loss	13 (20)	14 (56)	16 (25)	11 (44)
Positive family history of IBD	9/64 (14)	2 (8)	6/64 (9)	5 (20)
Extra-intestinal symptoms	4 (6)	9 (36)	6 (9)	7 (28)
Peri-anal lesions	9 (14)	4/24 (17)	7 (11)	6/24 (25)
≥1 alarm symptoms	38 (59)	24 (96)	39 (60)	23 (92)
Blood markers (n (%))				
haemoglobin (cut-off is age/sex specific[a])	5/61 (8)	6 (24)	4/61 (7)	7 (28)
C-reactive protein (> 10 mg/l)	5/56 (9)	5/20 (25)	2/53 (4)	8/23 (35)
erythrocyte sedimentation rate (> 20 mm/h)	8/59 (14)	8/24 (33)	4/59 (7)	12/24 (50)
Platelet count (> 450 x10^9/l)	4/61 (7)	3 (12)	4/61 (7)	3 (12)
≥1 blood marker	14/54 (26)	10/21 (40)	11/52 (21)	13/23 (56)
Faecal test (n (%))				
faecal calprotectin (> 50 µg/g)	14/63 (22)	13/22 (59)	9/63 (14)	18/22 (82)
Reference standard (n (%))				
endoscopy	9 (14)	20 (80)	4 (6)	25 (100)
Diagnoses (n (%))				
IBD	5 (8)	12 (48)	2 (3)	15 (60)
FGIDs	55 (85)	11 (44)	58 (89)	8 (32)

[a] 4-12 years < 7.1 mmol/l, boys 12-18 years < 8.1 mmol/l, girls 12-18 years < 7.4 mmol/l. Abbreviations: FGID: functional gastrointestinal disorder; IBD: inflammatory bowel disease

Table 3. The various diagnostic models for IBD of the non-imputed dataset and imputed dataset.

	Non-imputed DOR (95% CI)	Imputed Pooled DOR (95% CI)	Non-imputed AUC (95% CI)	Imputed Pooled AUC (95% CI)
Alarm symptoms	1.02 (1.01-1.04)	1.02 (1.01-1.04)	0.80	0.80
			(0.67-0.92)	(0.69-0.90)
Alarm symptoms +	1.02 (1.01-1.04)	1.02 (1.01-1.03)	0.88	0.85
C-reactive protein	1.18 (1.04-1.33)	1.14 (1.01-1.27)	(0.78-0.98)	(0.76-0.93)
Alarm symptoms +	1.01 (0.99-1.04)	1.02 (0.99-1.04)	0.97	0.97
faecal calprotectin	1.01 (1.003-1.02)	1.01 (1.003-1.03)	(0.93-1.00)	(0.93-1.00)

Basic model of alarm symptoms consists of the number of weighted alarm symptoms (Total score: 357): involuntary weight loss (weight: 44), rectal blood loss (weight: 60), family history of IBD (weight: 53), extra-intestinal symptoms (weight: 78), peri-anal lesions (weight: 71). The mean weighting scores for the alarm symptoms were based on the independent opinions of 85 physicians who treat children with chronic gastrointestinal symptoms from different clinical settings. The physicians weighted the alarm symptoms using a visual analogue scale from 0 (completely ruled out that the child had IBD) to 100 (absolutely sure that the child had IBD). Interpretation DOR: one point increase on a continuous scaled test result (weighted alarm symptoms, C-reactive protein, faecal calprotectin) increases the risk of IBD with the DOR-value. Abbreviations: DOR: Diagnostic Odds Ratio; AUC: area under the curve.

TEST CHARACTERISTICS

The median time intervals between blood sampling and endoscopy were 34 and 69 days for children with and without IBD, respectively. The median time interval between stool sampling and endoscopy was 4 days for children with IBD and 8 days for children without IBD. Of the 29 children who underwent endoscopy, 11 (2 missing values) had a delay of more than one month. Overall, 25 children had missing data for one or more variable of interest. The children with and without missing data were comparable in all baseline characteristics, except for extra-intestinal symptoms and setting (Appendix 2). To reduce selection bias, we therefore imputed the missing data.[26]

The diagnostic characteristics of alarm symptoms, blood markers, and faecal calprotectine are shown in Appendix 1. The AUC of faecal calprotectin was 0.98 (0.96–1.00), which was significantly higher than for all blood markers individually (p<0.05).

ADDED VALUE OF C-REACTIVE PROTEIN AND FAECAL CALPROTECTIN

The AUC of the basic model was 0.80 (0.69–0.90). Adding C-reactive protein or faecal calprotectin to the model increased the AUC to 0.85 (0.76–0.93) and 0.97 (0.93–1.00), respectively. However, this increase was only significant for faecal calprotectin (p<0.05) (Table 3). The goodness-of-fit test was good for all models (Hosmer and Lemeshow test p>0.05). The calibration plots of all models displayed linear concordance between the observed and predicted probabilities of IBD.

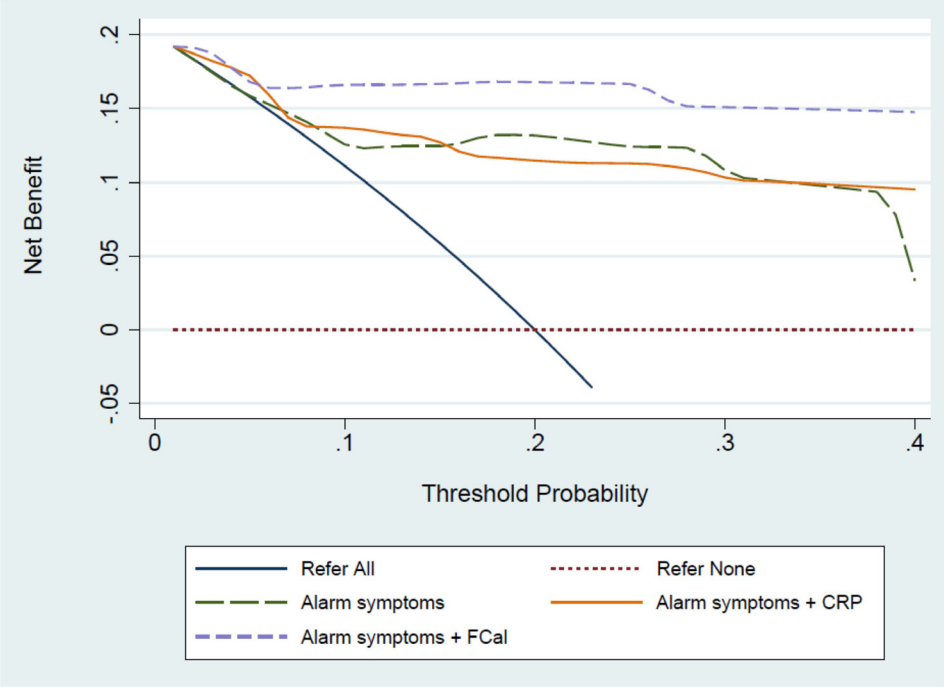


Figure 2. Decision curve for three models predicting the outcome of IBD of the non-imputed dataset. An example of interpreting the decision curve: the purple line representing the strategy of alarm symptoms + faecal calprotectin shows a net benefit of 0.16 at a threshold probability of 20%. The threshold probability of 20% means that a general practitioner would be willing to refer 5 children in order to prevent delay of diagnosis in 1 child with IBD. The net benefit of 0.16 means that this strategy would lead to the referral of 160 per 1000 children at risk, with all of these referrals positive for IBD. Abbreviations: CRP: C-reactive protein, FCal: faecal calprotectin.

Table 4. Reduction in referral for further diagnostic-work up per 100 children using different threshold probabilities for the three test strategies.

Threshold probability	Alarm symptoms	Alarm symptoms + C-reactive protein	Alarm symptoms + Faecal calprotectin
10%	-	15	32
15%	12	33	49
20%	35	40	59

Note: The reduction in the number of referral for further diagnostic-work up per 100 children without missing a child with IBD was calculated as follows: (net benefit strategy – net benefit of refer all) / [Pt / (1 – Pt)]*100.

DECISION CURVE ANALYSIS

The decision curve analysis indicated that all three diagnostic strategies had higher net benefit at diagnostic threshold probabilities of >5% when compared with the alternative of referring all children (Figure 2). Alarm symptoms in combination with faecal calprotectin had the highest net benefit at threshold probabilities between 5% and 40%. Table 4 shows the reduction in number of referrals without missing a child with IBD at different threshold probabilities for the three strategies.

DISCUSSION

In this study, 73% of children referred for evaluation of chronic gastrointestinal symptoms had functional gastrointestinal disorders that could be managed at primary care level. The prevalence of IBD was 19%. In children in whom the general practitioner considers referral for further diagnostic work-up, C-reactive protein provided no additional diagnostic value when used in combination with alarm symptoms. The addition of faecal calprotectin to alarm symptoms improved the AUC statistically significantly. The diagnostic test strategy of alarm symptoms and faecal calprotectin showed highest net benefit and reduced a higher number of referrals for IBD without a false negative compared to the strategy with alarm symptoms and C-reactive protein.

Physicians use a diagnostic threshold probability above which they initiate further testing before deciding on further management. The decision curve analysis indicates, that in children with a threshold probability of IBD <5%, testing with C-reactive protein or faecal calprotectin barely has added value. Therefore, we advise against testing C-reactive protein or faecal calprotectin in children with a very low risk for IBD (e.g. without alarm symptoms). At threshold probabilities between 5% and 40%, the test strategy with alarm symptoms and faecal calprotectin showed highest net benefit. Moreover, the discriminative power of faecal calprotectin proved to be superior to all blood markers, individually. Therefore, we can conclude that faecal calprotectin is the best laboratory test that is able to further stratify children who have already been identified as at risk for IBD by the classic assessment which includes history and physical examination. In our previous study we showed that in primary care a calprotectin value below 50 μg/g faeces can be used to safely rule out IBD.[30]

There are few publications about the added value of faecal calprotectin in children with symptoms suggestive of IBD. A study evaluating the added value of faecal calprotectin using the 'clinical eye' of the paediatrician showed that faecal calprotectin reduced the need for referral to a paediatric gastroenterologist, with only a low risk of missing a child with IBD.[31] However, in this study, we could not determine how this 'clinical eye' incorporated alarm symptoms and blood markers. Other researchers constructed a model to predict the probability of having IBD based on faecal calprotectin and age,[32] and this was used to correctly classify 85.5% of children with a sensitivity of 0.81 and specificity of 0.92 (AUC 0.92). However, important predictors, such as alarm symptoms and blood markers, were not included in the model.

It is important to realize that our study included patients referred by general practitioner

and general paediatrician. In these children, assumed to represent the same patient population, however, the children who were referred by a paediatrician displayed a higher IBD prevalence, and were older, had more alarm symptoms than children who were referred by their general practitioner. In a Dutch healthcare system this result is to be expected, since the paediatrician can only be consulted if they are referred by a general practitioner, and the paediatric gastroenterologist can be consulted if the children are referred by general practitioner or paediatrician. Healthcare systems in the United Kingdom, Scandinavia, Canada, New Zealand, and Australia are comparable.[33]

A limitation of our study is that alarm symptoms are routinely assessed with blood markers in children with symptoms suggestive of IBD. Consequently, the reference standard was interpreted with prior knowledge of the test results of alarm symptoms and blood markers, which might have caused review bias and overestimation of the diagnostic accuracy of the alarm symptoms and blood markers.[34] Another limitation is that we did not perform endoscopies in children with a low likelihood of IBD, because this was considered to be unethical. There are two important aspects to consider when assessing whether the use of two reference standards leads to biased estimates of accuracy: the verification pattern and the appropriateness of the follow-up.[35] Although the verification pattern was based on clinical judgement of the paediatric gastroenterologist and thereby to some degree subjective, our results showed that the children who received endoscopy at baseline were at higher risk for IBD than children who received follow-up. Therefore, if the test strategies were only evaluated in the children who received endoscopy at baseline the results were probably biased.[36] Because IBD is not a self-limiting disease and it is very rare that it stays in remission for one year,[31] there is a very small chance that we missed a child with IBD which alter the diagnostic value of tests. The use of a follow-up is the best achievable option given the reality of clinical care.[37]

The number of children with IBD in our study was low (17 children), so we combined the weighted alarm symptoms into one variable. The weighting was based on the opinion of 85 physicians and thereby subjective. However, the AUC of the combined alarm symptoms without weighting was comparable. Moreover, we evaluated the added value of two important markers, based on eight events per variable. The events per variable rule of thumb suggest you need at least 10 events per variable in the model, but five events per variable have been shown to be a valid generator for hypothesis.[38] A larger study with more events could develop a prediction model for IBD based on single alarm symptoms, blood markers, and faecal calprotectin.[24]

An evaluation of alarm symptoms and faecal calprotectin showed to be the optimal strategy for further stratifying children who have already been identified as at risk for IBD by the general practitioner. Notable is that this study focused on IBD alone, while general practitioners evaluate whether symptoms are related to any organic disease, including celiac disease. Further studies are needed to investigate whether the new test strategy of alarm symptoms in combination with faecal calprotectin actually improves decision-making. These studies should also evaluate costs and patients values.

REFERENCES

1. Peery AF, Dellon ES, Lund J, et al. Burden of gastrointestinal disease in the united states: 2012 update. *Gastroenterology*. 2012;143(5):1179-1187.
2. Spee LA, Lisman-Van Leeuwen Y, Benninga MA, Bierma-Zeinstra SM, Berger MY. Prevalence, characteristics, and management of childhood functional abdominal pain in general practice. *Scand J Prim Health Care*. 2013;31(4):197-202.
3. Kokkonen J, Haapalahti M, Tikkanen S, Karttunen R, Savilahti E. Gastrointestinal complaints and diagnosis in children: A population-based study. *Acta Paediatrica*. 2004;93(7):880-886.
4. Miele E, Simeone D, Marino A, et al. Functional gastrointestinal disorders in children: An italian prospective survey. *Pediatrics*. 2004;114(1):73-78.
5. Di Lorenzo C, Colletti RB, Lehmann HP, et al. Chronic abdominal pain in children: A technical report of the american academy of pediatrics and the north american society for pediatric gastroenterology, hepatology and nutrition: AAP subcommittee and NASPGHAN committee on chronic abdominal pain. *J Pediatr Gastroenterol Nutr*. 2005;40(3):249-261.
6. Rasquin A, Di Lorenzo C, Forbes D, et al. Childhood functional gastrointestinal disorders: Child/adolescent. *Gastroenterology*. 2006;130(5):1527-1537.
7. Gieteling MJ, Bierma-Zeinstra SM, Passchier J, Berger MY. Prognosis of chronic or recurrent abdominal pain in children. *J Pediatr Gastroenterol Nutr*. 2008;47(3):316-326.
8. van Tilburg MA, Venepalli N, Ulshen M, Freeman KL, Levy R, Whitehead WE. Parents' worries about recurrent abdominal pain in children. *Gastroenterology Nursing*. 2006;29(1):50-55.
9. Kim SC, Ferry GD. Inflammatory bowel diseases in pediatric and adolescent patients: Clinical, therapeutic, and psychosocial considerations. *Gastroenterology*. 2004;126(6):1550-1560.
10. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr*. 2014;58(6):795-806.
11. de Ridder L, Rings EH, Escher JC, CBO-werkgroep 'IBD bij kinderen'. Guideline 'diagnosis and treatment of inflammatory bowel disease in children'. *Ned Tijdschr Geneeskd*. 2010;154:A1898.
12. Holtman GA, Lisman-van Leeuwen Y, Reitsma JB, Berger MY. Noninvasive tests for inflammatory bowel disease: A meta-analysis. *Pediatrics*. 2016;137(1):1-11.
13. Vermeire S, Van Assche G, Rutgeerts P. Laboratory markers in IBD: Useful, magic, or unnecessary toys? *Gut*. 2006;55(3):426-431.
14. van Rheenen PF, Van de Vijver E, Fidler V. Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: Diagnostic meta-analysis. *BMJ*. 2010;341:c3369.
15. Henderson P, Anderson NH, Wilson DC. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease: A systematic review and meta-analysis. *Am J Gastroenterol*. 2014;109(5):637-645.
16. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers and diagnostic tests. *BMJ*. 2016;352:i6.
17. Holtman GA, Lisman-van Leeuwen Y, Kollen BJ, et al. Challenges in diagnostic accuracy studies in primary care: The fecal calprotectin example. *BMC Fam Pract*. 2013;14:179-2296-14-179.
18. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol*. 1997;32(9):920-924.
19. Oord T, Hornung N. Fecal calprotectin in healthy children. *Scand J Clin Lab Invest*. 2014;74(3):254-258.

6

20. Gisbert J, McNicholl A. Questions and answers on the role of faecal calprotectin as a biological marker in inflammatory bowel disease. *Digestive and Liver Disease*. 2009;41(1):56-66.
21. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49(12):1373-1379.
22. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148(3):839-843.
23. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574.
24. Moons KG, de Groot JA, Linnet K, Reitsma JB, Bossuyt PM. Quantifying the added value of a diagnostic test or marker. *Clin Chem*. 2012;58(10):1408-1417.
25. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138.
26. van der Heijden, Geert JMG, Donders ART, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. *J Clin Epidemiol*. 2006;59(10):1102-1109.
27. Vink G, Frank LE, Pannekoek J, Buuren S. Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*. 2014;68(1):61-90.
28. Moons KG, Donders RA, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59(10):1092-1101.
29. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines. *BMC Med Res Methodol*. 2009;9:57-2288-9-57.
30. Holtman G.A., Lisman-van Leeuwen Y., Kollen B.J., et al. Diagnostic accuracy of faecal calprotectin for paediatric inflammatory bowel disease in primary care; a prospective cohort study. *Ann Fam Med* 2016. doi: 10.1370/afm.1949.
31. Van de Vijver E, Schreuder AB, Cnossen WR, Muller Kobold AC, van Rheenen PF, North Netherlands Pediatric IBD Consortium. Safely ruling out inflammatory bowel disease in children and teenagers without referral for endoscopy. *Arch Dis Child*. 2012;97(12):1014-1018.
32. Degraeuwe PL, Beld MP, Ashorn M, et al. Faecal calprotectin in suspected paediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2015;60(3):339-346.
33. Thomson S, Osborn R, Squires D, Reed SJ. International profiles of health care systems 2011: The Commonwealth Fund. http://www.commonwealthfund.org/publications/fund-reports/2011/nov/international-profiles-of-health-care-systems-2011. Published Nov 2011.
34. Oostenbrink R, Moons KG, Bleeker SE, Moll HA, Grobbee DE. Diagnostic research on routine care data: Prospects and problems. *J Clin Epidemiol.* 2003;56(6):501-506.
35. Naaktgeboren CA, De Groot JA, van Smeden M, Moons KG, Reitsma JB. Evaluating diagnostic accuracy in the face of multiple reference standards. *Ann Intern Med*. 2013;159(3):195-202.
36. Naaktgeboren CA, de Groot JA, Rutjes AW, Bossuyt PM, Reitsma JB, Moons KG. Anticipating missing reference standard data when planning diagnostic accuracy studies. *BMJ*. 2016;352:i402.
37. Knottnerus JA, Dinant GJ. Medicine based evidence, a prerequisite for evidence based medicine. *BMJ*. 1997;315(7116):1109-1110.
38. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and cox regression. *Am J Epidemiol*. 2007;165(6):710-718.

39. Pels LP, Van de Vijver E, Waalkens HJ, et al. Slow hematological recovery in children with IBD-associated anemia in cases of "expectant management". *J Pediatr Gastroenterol Nutr*. 2010;51(6):708-713.
40. Mack DR, Langton C, Markowitz J, et al. Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics*. 2007;119(6):1113-1119.

6

Appendix 1. Diagnostic characteristics of alarm symptoms, blood markers and faecal calprotectin for IBD with a pre-test probability of 19%, using the imputed dataset.

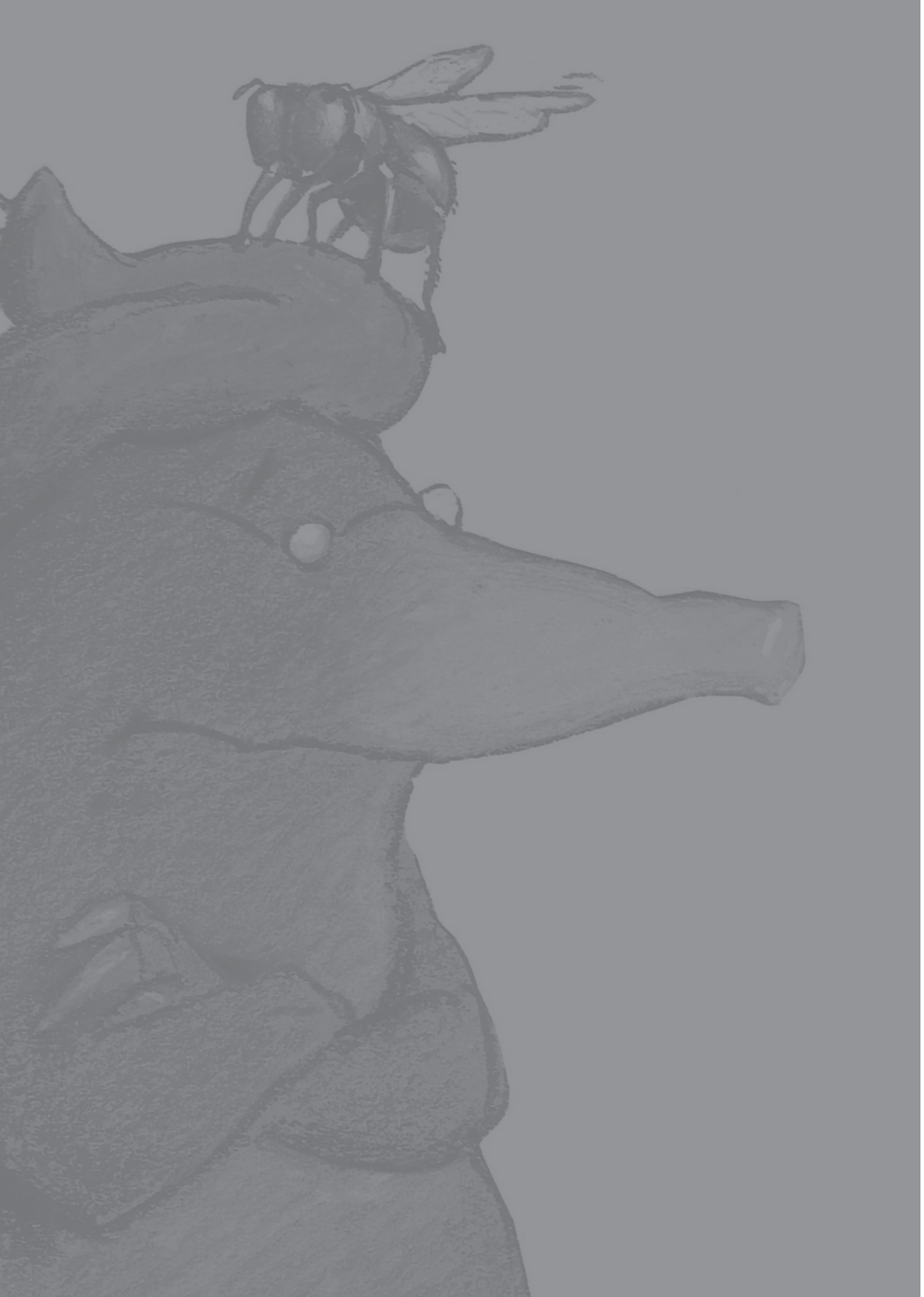
	Sens (95% CI)	Spec (95% CI)	PPV (95% CI)	NPV (95% CI)	LR+ (95% CI)	LR – (95% CI)	DOR (95% CI)	Area under ROC curve (95% CI)
Alarm symptoms								
Involuntary weight loss	0.69 (0.45-0.85)	0.85 (0.75-0.91)	0.52 (0.33-0.71)	0.92 (0.83-0.96)	4.53 (2.41-8.50)	0.37 (0.18-0.75)	12.22 (3.70-40.40)	-
Rectal blood loss	0.49 (0.28-0.70)	0.74 (0.63-0.83)	0.31 (0.17-0.50)	0.86 (0.75-0.92)	1.90 (1.01-3.53)	0.69 (0.43-1.11)	2.74 (0.94-8.04)	-
Positive family history of IBD	0.11 (0.03-0.34)	0.87 (0.78-0.93)	0.18 (0.51-0.47)	0.80 (0.70-0.88)	0.91 (0.22-3.84)	1.01 (0.84-1.22)	0.90 (0.18-4.59)	-
Growth failure	0.00 (0.00-0.18)	0.92 (0.83-0.96)	0.00 (0.00-0.39)	0.79 (0.69-0.87)	0.00 (-)	1.09 (1.02-1.17)	0.00 (-)	-
Extra-intestinal symptoms	0.29 (0.13-0.52)	0.89 (0.79-0.94)	0.38 (0.18-0.65)	0.84 (0.74-0.90)	2.59 (0.97-6.96)	0.80 (0.59-1.09)	3.23 (0.91-11.51)	-
Peri-anal lesions	0.39 (0.20-0.62)	0.90 (0.81-0.95)	0.50 (0.26-0.73)	0.86 (0.77-0.92)	4.09 (1.64-10.19)	0.67 (0.46-0.99)	6.10 (1.77-20.99)	-
Blood markers								
C-reactive protein	0.50 (0.29-0.71)	0.95 (0.87-0.98)	0.69 (0.42-0.87)	0.89 (0.80-0.94)	9.4 (3.2-27.3)	0.53 (0.53-0.84)	17.8 (4.46-71.20)	0.79 (0.69-0.90)
erythrocyte sedimentation rate	0.57 (0.35-0.77)	0.91 (0.82-0.96)	0.60 (0.37-0.79)	0.90 (0.81-0.95)	6.19 (2.71-14.17)	0.47 (0.27-0.82)	13.11 (3.81-45.14)	0.80 (0.68-0.92)
Platelet count	0.18 (0.07-0.41)	0.94 (0.86-0.98)	0.43 (0.16-0.74)	0.83 (0.73-0.89)	3.09 (0.80-11.90)	0.87 (0.69-1.09)	3.55 (0.75-16.83)	0.70 (0.57-0.84)
haemoglobin	0.43 (0.24-0.66)	0.95 (0.88-0.98)	0.68 (0.40-0.88)	0.88 (0.78-0.93)	9.01 (2.84-28.55)	0.59 (0.39-0.90)	15.16 (3.63-63.37)	0.77 (0.65-0.88)
Faecal marker								
faecal calprotectin	0.99 (0.81-1.00)	0.84 (0.74-0.91)	0.60 (0.42-0.75)	1.00 (0.94-1.00)	6.17 (3.65-10.44)	0.00 (0.00-3.33)	-	0.98 (0.96-1.00)

Cut-offs blood markers and faecal calprotectin: haemoglobin (4-12 years < 7.1 mmol/l, boys 12-18 years < 8.1 mmol/l, girls 12-18 years < 7.4 mmol/l), C-reactive protein (> 10 mg/l), erythrocyte sedimentation rate (> 20 mm/h), Platelet count (> 450 x10^9/l), faecal calprotectin (> 50 µg/g). Abbreviations: CI: Confidence Interval; Sens: sensitivity; Spec: Specificity; NPV: negative predictive value; PPV: positive predictive value; LR+: positive likelihood ratio; DOR: Diagnostic Odds Ratio; ROC: receiver-operator curve.

Appendix 2. Complete data, number of missing per variable and difference in distribution between children with and without missing data.

	Complete data N = 65 (72%)	Missing N (%)	P-value
Male sex (n (%))	28 (43.1)	0	0.8
Age in years at baseline (median, IQR)	11 (8-15)	0	0.6
IBD (n (%))	13 (20)	1 (1.1)	0.4
Setting (n (%))		0	0.003*
Primary care	22 (33.8)		
Secondary care	27 (41.5)		
Tertiary care	16 (24.6)		
Alarm symptoms (n (%))			
Growth failure	5 (7.7)	0	1.0
Involuntary weight loss	16 (24.6)	0	0.6
Rectal blood loss	21 (32.3)	0	0.6
Positive family history of IBD	8 (12.3)	1 (1.1)	0.4
Extra-intestinal symptoms	7 (10.8)	0	0.03*
Peri-anal lesions	11 (16.9)	1 (1.1)	0.2
Blood markers (median (IQR))			
haemoglobin (mmol/l)	8 (7.7-8.4)	5 (5.6)	0.8
C-reactive protein (mg/l)	1 (1-3.6)	15 (16.7)	0.8
erythrocyte sedimentation rate (mm/h)	7 (4-13.5)	8 (7.8)	0.4
Platelet count (x10^9/l)	299 (253-359)	5 (5.6)	0.6
Faecal test (median (IQR))			
Faecal calprotectin (µg/g)	22 (20-101)	5 (5.6)	0.6

*P < 0.05; Difference between children with and without missing data was tested with chi^2 for categorical variables, Independent sample T-Test for continuous scaled normally distributed variables, Mann-Whitney test for continuous scaled non-normally distributed variables.



CHAPTER 7

EVALUATION OF POINT-OF-CARE TEST CALPROTECTIN AND LACTOFERRIN FOR INFLAMMATORY BOWEL DISEASE AMONG CHILDREN WITH CHRONIC GASTROINTESTINAL SYMPTOMS

Gea A Holtman

Yvonne Lisman-van Leeuwen

Patrick F van Rheenen

Boudewijn J Kollen

Johanna C Escher

Angelika Kindermann

Yolanda B de Rijke

Marjolein Y Berger

Family Practice; in press

ABSTRACT

BACKGROUND

Faecal calprotectin is considered to be a valid test for ruling out inflammatory bowel disease (IBD) in children with chronic gastrointestinal symptoms in specialist care. In contrast, faecal lactoferrin has higher specificity. The recent availability of both as point-of-care tests (POCTs) makes them attractive for use in primary care.

OBJECTIVE

To evaluate the test characteristics of calprotectin and lactoferrin POCTs for diagnosing IBD in symptomatic children.

METHODS

We defined two prospective cohorts of children with chronic gastrointestinal symptoms: 1) children presenting to primary care (primary care cohort); 2) children referred for specialist care (referred cohort). Baseline POCT results were compared with the outcome of either endoscopic assessment or 12 months follow-up. Clinicians were blinded to the POCT results.

RESULTS

In the primary care cohort, none of the 114 children had IBD, and the calprotectin and lactoferrin POCTs had specificities of 0.95 (0.89–0.98) and 0.98 (0.93–0.99), respectively. In the referred cohort, 17 of the 90 children had IBD: the sensitivity of POCT calprotectin and POCT lactoferrin were both 0.94 (0.72-0.99); and the specificity was 0.93 (0.84-0.97) and 0.99 (0.92-1.00), respectively. The POCT calprotectin could reduce the referral rate by 76% and POCT lactoferrin by 81%, while missing one child with IBD (6%).

CONCLUSION

A diagnostic test strategy in primary care using a simple POCT calprotectin or lactoferrin has the potential to reduce the need for referral for further diagnostic work-up in specialist care, with a low risk of missing a child with IBD.

INTRODUCTION

In primary care, it can be difficult to differentiate between functional gastrointestinal disorders and organic disease, such as inflammatory bowel disease (IBD). Indeed, specialist care referral can ensure early diagnosis and treatment of children with IBD, and thereby reduce complications.[1] Typically, referral is recommended for children with red flags, such as rectal bleeding, weight loss, or a family history of IBD.[2] These red flags, however, are common and have little discriminative power.[3,4] Therefore, simple, accurate tests are needed that can differentiate between functional and organic disorders in children with chronic gastrointestinal symptoms.

Faecal calprotectin, a non-invasive test for intestinal inflammation, has a high rule-out value among symptomatic children in specialist care (high sensitivity, reduced post-test probability of a normal test result).[5-7] We have also shown a high rule-out value in children presenting in primary care.[8] In both settings, faecal calprotectin was measured by enzyme-linked immunosorbent assays (ELISA); however, point-of-care tests (POCT) have been developed with comparable accuracy to the ELISA.[9] The concentration of calprotectin is stable at room temperature for 7 days, and only a few grams of stool are required. This enables the patient to deliver a stool sample at the general practice and receive a test result within 15 minutes. Whereas faecal calprotectin, with its high sensitivity, is the most used faecal diagnostic test in children suspected for IBD, faecal lactoferrin has been shown to have a high specificity for IBD.[10,11] Thus, lactoferrin might be of additional diagnostic value when used in combination with faecal calprotectin in point-of-care testing.

To date, the diagnostic accuracy of the calprotectin POCT, lactoferrin POCT and the combination of both, has not been examined for IBD among children with chronic gastrointestinal symptoms in primary care. Therefore, we studied the test characteristics in children presenting in primary care and children referred to specialist care.

METHODS

DESIGN AND PATIENTS

This was a prospective cohort study with a delayed-type cross-sectional design.[12] Children were included in the Netherlands from July 2011 to July 2013 and were followed for 12 months. Parents of all children, and children aged ≥12 years, provided written informed consent, as appropriate. The study protocol was approved by the medical ethics committee of the University Medical Centre Groningen.

We included two cohorts of children with chronic gastrointestinal symptoms. The primary care cohort included consecutive children recruited by 64 participating general practitioners (GPs) from 38 practices. The referred cohort included children selected from the primary care cohort who were referred for specialist care based on the presence of ≥1 red flags, and consecutive children referred by GPs and paediatricians who were included at one of four general hospitals and three academic centres.

The inclusion criteria were age 4–18 years, chronic diarrhoea (score 5, 6 or 7 on the

Table 1. Definitions of red flags for inflammatory bowel disease.

Red flags	Measurement	Positive
Involuntary weight loss	History	Involuntary decrease in weight of > 1 kg
Rectal bleeding	History	Rectal bleeding with defecation without constipation according to ROME III criteria
Family history of IBD	History	Affected first-degree relative(s)
Growth failure	History and physical examination	Target height range > -1 standard deviation score
Extra-intestinal symptoms	Physical examination	Eyes (episcleritis, scleritis, uveitis), skin (erythema nodosum, pyoderma gangrenosum, psoriasis), mouth ulcers, finger clubbing, arthritis
Peri-anal lesions	Physical examination	Skin tag, haemorrhoid, fissure, fistula or abscess
Hemoglobin	Local laboratory	4-12 years <7.1 mmol/l, boys 12-18 years <8.1 mmol/l, girls 12-18 years <7.4 mmol/l[22]
C-reactive protein	Local laboratory	>10 mg/l[4]
Erythrocyte sedimentation rate	Local laboratory	>20 mm/h[4]
Platelet count	Local laboratory	>450 x10^9/l[23]

IBD: inflammatory bowel disease

Bristol Stool Form Scale[13] for ≥2 weeks or ≥2 episodes in the past 6 months) and/or recurrent abdominal pain (≥2 episodes of abdominal pain or discomfort in the past 6 months). Exclusion criteria were known chronic organic gastrointestinal disease; endoscopic evaluation or faecal calprotectin testing in the preceding 6 months; chronic use of antibiotics, non-steroid anti-inflammatory drugs or oral corticosteroids in the previous 6 months; or difficulties in understanding questionnaires.

PATIENT FLOW

At baseline, the study GP or paediatrician assessed the presence of involuntary weight loss, rectal bleeding, family history of IBD, growth failure, extra-intestinal symptoms and peri-anal lesions, based on a structured history and physical exam. In addition, a blood sample (for assessment of haemoglobin, C-reactive protein, erythrocyte sedimentation rate and platelet count) was taken (see red flags, Table 1). All children were clinically followed for 12 months. Children with ≥1 red flags (Table 1) at baseline or after 12 months' follow-up were referred to a paediatric gastroenterologist who decided whether the child required endoscopic evaluation based on global assessment, physical examination, and blood test results. Children who did not had an indication for endoscopy received the reference standard of a clinical 12 months follow-up, because it was thought to be unethical to perform an invasive procedure under

full anaesthesia in children with a low likelihood of IBD. The GPs or paediatricians of children lost to follow-up were contacted after 12 months in order to receive the most recent information required for making the final diagnosis.

POINT-OF-CARE TESTING

A stool sample tube, together with information on how to collect a stool sample, was provided at baseline. Stool was collected at home, shortly after inclusion, and sent to a laboratory for storage at -80°C until data collection was complete (September 2014). We used the dual calprotectin–lactoferrin POCT (CerTest Calprotectin–Lactoferrin combo, Zaragoza, Spain), which is a semi-quantitative immunochromatographic assay with a visual reading device. The test was performed at the laboratory of the Erasmus University Medical Centre in Rotterdam, according to the manufacturer's instructions.

The stick of the stool collection tube was dipped into different parts of the stool sample and added to the extraction buffer in the stool collection tube; or, for liquid samples, 15 µl was pipetted into the stool collection tube. The tube was shaken to ensure good sample distribution. Four drops from the sample in the stool collection tube were added to both windows of the POCT device (Window A for calprotectin and Window B for lactoferrin) and the results were read after 10 minutes. If a red-coloured line appeared in either window, the respective test was positive. The thresholds for positivity were >50 µg/g faeces and >10 µg/g faeces for the calprotectin and lactoferrin tests, respectively. The results were invalid when the green control line was absent. The testers were blinded to the clinical characteristics and diagnosis, and all clinicians and researchers were blinded to the outcomes of both POCTs.

DIAGNOSIS

IBD was diagnosed by esophagogastroduodenoscopy and ileocolonoscopy with histopathology of multiple biopsies, according to the revised Porto Criteria.[14] Absence of IBD was defined as no endoscopic and histopathological evidence of IBD and/or no indication for endoscopy within or at 12 months' follow-up.

STATISTICAL ANALYSIS

A simple and non-invasive POCT might convince the GP to use this test in children with a low likelihood of IBD. In this population, false-positive results should be minimized to avoid unnecessary referrals for endoscopy. Therefore, we were mostly interested in a precise estimate of the specificity of both POCT results in the primary care cohort. A sample size calculation, which was based on a IBD prevalence rate of 5%, loss to follow-up of 10%, and specificity of 93%, generated a total of 118 children to determine the specificity with 95% confidence intervals (CIs) and adequate precision (i.e. 5%).[12] In children in whom the GP considers a referral the likelihood of IBD increases. In this population, low false-negative rates are important to prevent delay in the diagnoses of children with IBD. Therefore, we were mostly interested in a precise estimate of sensitivity of both POCTs in the referred cohort. Based on a IBD prevalence rate of 20%, loss to follow-up of 10% and sensitivity of 95%, we calculated a sample size of 100 children to determine sensitivity with 95% CIs and adequate precision (i.e. 10%).[12] Moreover, we calculated specificity, positive predictive value

(PPV), negative predictive value (NPV) with 95% CIs separately for each POCT in the referred cohort. In order to evaluate spectrum bias, subgroup analyses in children who were referred 1) by GP and 2) by general paediatrician were performed. To evaluate the accuracy of the structured clinical assessment in this study, we compared the proportion of children with red flags and the proportion of children with IBD in children with and without an endoscopy at baseline. Moreover, we evaluated the potential clinical impact of the different test results on referral rate or missed IBD diagnosis of the dual calprotectin-lactoferrin POCT with scenario analysis in the primary care cohort and referred cohort. We performed complete case analyses. Statistical analyses were performed with IBM SPSS for Windows, Version 20.0 (IBM Corp., Armonk, NY, USA).

RESULTS

DESCRIPTIVE CHARACTERISTICS

The participant flow and baseline characteristics are summarized in Figure 1 and Table 2. We included 114 and 90 children in the primary care cohort and referred cohort, respectively. None of the primary care cohort and 17 (19%) of the referred cohort were diagnosed with IBD (Appendix 1). In the referred cohort, children referred by a general paediatrician more frequently had involuntary weight loss, rectal bleeding and extra-intestinal symptoms, and were more likely to be diagnosed with IBD, when compared with the children referred by a GP. Children who were subjected to an endoscopy at baseline had more frequently red flags (100%) and high risk for IBD (60%) than children who underwent no endoscopy at baseline (66% and 3%, respectively) (Table 2).

The median intervals between faecal sampling and endoscopy were 4 and 8 days for children with and without IBD, respectively; however, 11 of the 27 children (2 missing samples) who underwent endoscopy had a delay >1 month. POCTs were not performed in 12 children, and 9 test results (7.9%) were missing in the primary care cohort and 5 (5.6%) in the referred cohort.

THE POCT CHARACTERISTICS

The characteristics of the calprotectin and lactoferrin POCTs for IBD, in children in primary care cohort and referred cohort are detailed in Table 3. In the primary care cohort, specificity of POCT calprotectin and POCT lactoferrin were 0.95 (0.89-0.98) and 0.98 (0.93-0.99), respectively. In the referred cohort, sensitivity of POCT calprotectin and POCT lactoferrin were both 0.94 (0.72-0.99); and the specificity was 0.93 (0.84-0.97) and 0.99 (0.92-1.00), respectively. In a subgroup analysis, the sensitivity of both POCTs in children referred to specialist care by their GP was 1.00 (0.57-1.00), the specificity for POCT calprotectin was 0.93 (0.84-0.97) and for POCT lactoferrin 1.00 (0.94-1.00). The sensitivity and specificity of both POCTs in children referred to specialist care by their general paediatrician were 0.91 (0.62-0.93).

SCENARIO ANALYSIS

Table 4 shows that in primary care cohort 97% of the children had the same outcome for

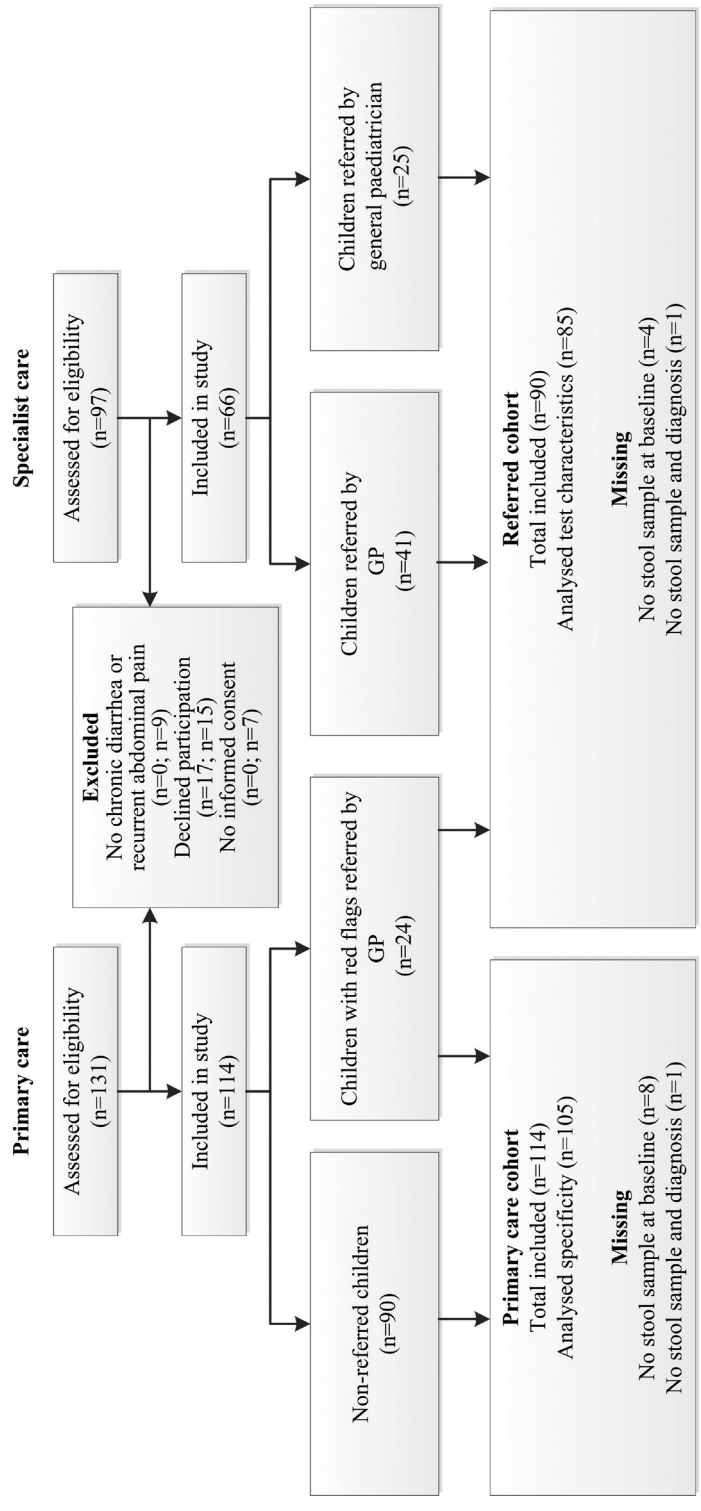


Figure 1. Patient flow in the study.
We included 114 children in the primary care cohort, of whom 24 were referred to specialist care based on ≥1 red flags, 2 underwent endoscopy at baseline and 112 were followed for 12 months. We included 90 children in the referred cohort, of whom 65 (72%) were referred by a GP and 25 (28%) were referred by a general paediatrician. At baseline, 25 children underwent endoscopic and histopathological evaluation, and during follow-up, 4 children underwent endoscopy. The remaining 61 children had no indication for endoscopy during the 12 months' follow-up. Children who were primarily seen in primary care, but who were referred to specialist care, were evaluated in both analyses. Point-of-care testing was not performed in 12 children because faecal samples were not collected (n = 9), were not stored at the laboratory (n = 2) or were insufficient for analysis (n = 1).

Table 2. Baseline characteristics, baseline red flags, number of endoscopies, and number and type of inflammatory bowel disease in the primary care cohort (n=114) and referred cohort (n=90).

	Main analysis		Subgroups referred cohort		Subgroups referred cohort	
	Primary care cohort (n = 114)	**Referred cohort (n = 90)**	**Referred by GP[b] (n = 65)**	**Referred by general paediatrician (n = 25)**	**No endoscopy at baseline (n = 65)**	**Endoscopy at baseline (n = 25)**
Male (n (%))	38 (33)	37 (41)	29 (45)	8 (32)	27 (42)	10 (40)
Age at baseline (median, IQR)	9 (6-12)	11 (7-15)	10 (7-14)	14 (10-15.5)	9 (6-14)	15 (12-16)
Red flags (n (%))						
- Involuntary weight loss	5 (4)	23 (26)	10 (15)	13 (52)	10 (15)	13 (52)
- Rectal bleeding	7 (6)	27 (30)	13 (20)	14 (56)	16 (25)	11 (44)
- Family history of IBD	5 (4)	11 (12)	9/64 (14)	2 (8)	6/64 (9)	5 (20)
- Growth failure	4 (3)	6 (7)	6 (9)	0 (0)	5 (8)	1 (17)
- Extra-intestinal symptoms	0 (0)	13 (14)	4 (6)	9 (36)	6 (9)	7 (28)
- Peri-anal lesions	7 (6)	13 (14)	9 (14)	4/24 (17)	7 (11)	6/24 (25)
- Hb (age/sex specific[a])	1/111 (1)	11/86 (13)	5/61 (8)	6 (24)	4/61 (7)	7 (28)
- CRP (>10 mg/l)	2/110 (2)	10/76 (13)	5/56 (9)	5/20 (25)	2/53 (4)	8/23 (35)
- ESR (>20 mm/h)	5/111 (5)	16/83 (19)	8/59 (14)	8/24 (33)	4/59 (7)	12/24 (50)
- Platelet count (>450 x10^9/l)	4/111 (4)	7/86 (8)	4/61 (7)	3 (12)	4/61 (7)	3 (12)
Red flags (n (%))	29 (25)	68 (76)	43 (66)	25 (100)	43 (66)	25 (100)
Endoscopy (n (%))	2 (2)	29 (32)	9 (14)	20 (80)	4 (6)	25 (100)
IBD (n (%))	0	17 (19)	5/64 (8)	12 (48)	2 (3)	15 (60)
- Crohn's disease		7	3	4	0	7
- ulcerative colitis		8	2	6	2	6
- IBD-unclassified		2	0	2	0	2

Red flags: growth failure, involuntary weight loss, rectal bleeding, family history of IBD, extra-intestinal symptoms, peri-anal lesions, positive blood markers (Hb, CRP, ESR, platelet count).

Abbreviations: IBD, inflammatory bowel disease; GP, general practitioner; Hb, haemoglobin; ESR, erythrocyte sedimentation rate; CRP, C-reactive protein; IQR, interquartile range

[a] 4-12 years <7.1 mmol/l, boys 12-18 years <8.1 mmol/l, girls 12-18 years <7.4 mmol/l

[b] Children were also referred by GPs who did not participate in this study.

Table 3. Test characteristics of the calprotectin and lactoferrin POCTs in primary care cohort and referred cohort.

	Sens (95%CI)	**Spec (95% CI)**	**PPV (95% CI)**	**NPV (95% CI)**
Primary care cohort (n=105)				
FCal POCT	-	0.95 (0.89-0.98)	-	-
FLacto POCT	-	0.98 (0.93-0.99)	-	-
Referred cohort (n=85)				
FCal POCT	0.94 (0.72-0.99)	0.93 (0.84-0.97)	0.75 (0.53-0.89)	0.98 (0.92-1.00)
FLacto POCT	0.94 (0.72-0.99)	0.99 (0.92-1.00)	0.94 (0.72-0.99)	0.99 (0.92-1.00)

Abbreviations: FCal: faecal calprotectin; FLacto: faecal lactoferrin; POCT, point-of-care test; PPV, positive predictive value; NPV, negative predictive value; Sens, sensitivity; Spec, specificity.

Table 4. Scenario analysis of the POCT combination results in both cohorts.

Primary care cohort (n=105)		**Children with red flags (n=22)**	**Children without red flags (n=83)**
Test strategies	**N (%)**	**Reduction of referrals**	**Referrals**
FCal - FLacto -	100 (95%)	21 (95%)	-
FCal - FLacto +	0	-	-
FCal + FLacto -	3 (3%)	1 (5%)	2 (2%)
FCal + FLacto +	2 (2%)	-	2 (2%)
Referred cohort (n=85)		**Referrals (n=85)**	**IBD (n=16)**
POCTs results indicates no referral			
Test strategies	**N (%)**	**Reduction of referrals**	**Missed IBD cases**
FCal - FLacto -	65 (76%)	65 (76%)	1 (6%)
FCal + FLacto -	4 (5%)	4 (5%)	0
FCal - FLacto +	0	-	-
POCTs results indicates referral			
Test strategies	**N (%)**	**Referrals**	**IBD**
FCal + Flacto +	16 (19%)	16 (19%)	15 (94%)
FCal + FLacto -	4 (5%)	4 (5%)	0
FCal - Flacto +	0	-	-

Abbreviations: FCal, faecal calprotectin; FLacto, faecal lactoferrin; POCT, point-of-care test; +, positive POCT; -, normal POCT

Notes: None of the children in primary care cohort were diagnosed with IBD. The test strategies are based on the assumption that children with both normal POCTs results should not be referred and both positive POCTs results should be referred. We showed both the impact on referrals and reduction of referrals for a combination of positive and normal result. Missing POCT: two children with red flags in primary care cohort; seven children without red flags in primary care cohort; five children in referred cohort; one child with IBD.

both POCTs and in referred cohort 95%. In the primary care cohort, a normal POCT calprotectin and lactoferrin could hypothetically reduce 95% of referrals of children with red flags compared to referral of all children with red flags without missing a child with IBD. In the referred cohort, a normal POCT calprotectin and lactoferrin could hypothetically reduce referrals by 76%, compared to referral of all children considered for referral, but with the risk that 1 child (6%) with IBD would be missed (Table 4).

DISCUSSION

SUMMARY

The POCTs showed high specificities in children presenting with chronic gastrointestinal symptoms in primary care, none of which were diagnosed with IBD. In children referred for further diagnostic work-up, both POCTs showed high sensitivities and negative predictive values. In both cohorts 95% or more of the children had the same outcome for both POCTs, which indicates that the tests provided little additional value to one another.

STRENGTHS AND LIMITATIONS

Evaluating test characteristics is a challenge when the prevalence of a disease is very low: a case-control design would overestimate test characteristics, while the preferred method of including consecutive at-risk children would be extremely time consuming and costly. In addition, using an invasive reference standard, risks exposing healthy children unnecessarily and would be ethically intolerable. Nevertheless, test characteristics urgently need to be evaluated in primary care.[15] Based on this need, we used a pragmatic design that included consecutive children presenting with chronic gastrointestinal symptoms in primary care and selected high-risk children who were referred for specialist care.[12] As expected, the prevalence and severity of IBD differed between these groups. In the referred cohort, children referred by a general paediatrician were more severely ill than children referred by their GP. This finding might reflect that, in the Netherlands, a paediatric gastroenterologist can see children if they are directly referred by a GP or indirectly referred by a paediatrician. Comparable healthcare systems exist in the United Kingdom, Scandinavia, Canada, New Zealand and Australia.[16] The sensitivity and specificity of both POCTs was slightly lower in the children who were referred by their general paediatrician than those who were referred by their GPs. Therefore, when interpreting our results, the effect of spectrum bias should be taken into account.

As reference standards for diagnosis of IBD, we used endoscopy and 12 months' follow-up, because these are consistent with clinical experience. This may have introduced bias.[17] Nevertheless we feel the risk of misclassifying a child with IBD was extremely low with the introduction of follow-up, given that IBD is a chronic disease that often becomes clinically manifest within 12 months of presentation. Thus, children without red flags or a clinical indication for endoscopy over 12 months are very likely not to have IBD. In addition, IBD might have been missed at endoscopy, but persisting symptoms during follow-up may lead to further evaluation, therewith the risk of missing a child with IBD further decreases. Although the decision of the paediatric gastroenterologist was based on structured clinical information, the clinical judgement is to some degree subjective. The triage seemed efficient, because the children who were subjected to an endoscopy at baseline had more frequently red flags and higher risk for IBD than children who received no endoscopy at baseline. Although the use of two reference standards is not the ideal situation, following children for 12 months is the best option and represents reality of clinical care.[18]

A GP or paediatrician would only have immediate access to a test result of calprotectine or lactoferrin if all children with abdominal symptoms bring along a sample of faeces. This is not feasible nor desirable. 'Point-of-care' therefore needs some explanation. In daily practice, the diagnostic strategy incorporates two consultations, one in which the test is ordered and a second in which the results are discussed. It is the latter consultation that will increase efficiency because the POCT can be conducted by a GP assistant, results will be available within 15 minutes and the patient can be seen immediately by a GP to discuss the results.

COMPARISON WITH EXISTING LITERATURE

We found no previous studies that have evaluated the characteristics of POCTs for discriminating IBD from other gastrointestinal disorders in children with chronic gastrointestinal symptoms. In one study, both tests were measured by ELISA and were significantly elevated in children with active IBD.[19] The interaction between both tests was taken to indicate that the tests should be used together. We also expected that calprotectin and lactoferrin had added value to one another, not least because calprotectin showed high sensitivity and lactoferrin showed high specificity.[5,10,11] Although lactoferrin had higher specificity than calprotectin, the improvement in diagnostic value of using a combination POCT was not substantial to both tests individually.

In a primary care study of adults with lower gastro-intestinal abdominal symptoms, researchers evaluated the diagnostic performance of both POCTs for identifying IBD.[20] Although they produced comparable specificities to our results, their sensitivities were lower. This might be explained by the adult study population and the use of different test assays and cut-off values.

IMPLICATIONS FOR RESEARCH AND PRACTICE

No physician wants to miss a diagnosis of IBD in children presenting with chronic gastrointestinal symptoms; likewise, he or she must minimize referrals of children without IBD. In primary care, IBD has a low prevalence, so the probability of missing a child with IBD will be small and the probability of a referral of a child with functional gastrointestinal disorder will be high. To reduce the referral rate a test must have high specificity, which both POCTs (especially POCT lactoferrin) showed in the primary care cohort. The low false positive rates have the potential to reduce the number of referrals of children without IBD. The scenario-analysis showed that a normal POCT calprotectin or POCT lactoferrin test in children with red flags could reduce the number of referrals with 95% or 100%, respectively. None of the children without red flags was diagnosed with IBD during the 12 months follow-up. Therefore, testing with both POCTs seems not of value in children without red flags. A disadvantage is that our sample was too small to include sufficient children with IBD in the primary care cohort. Therefore, we could not present positive and negative predictive values.

7

High sensitivities and negative predictive values are needed to minimize the rate of negative test results in children with IBD, which in turn, can delay further testing and treatment.[21] The high sensitivities and negative predictive values of both POCTs in the referred cohort indicate that it can effectively rule out IBD in children in whom the GP considers referral. In addition, we showed that it could reduce the referral rate by 76% (POCT calprotectin) and 81% (POCT lactoferrin), while missing one child with IBD (6%). What is notable is that this study did not take into account referrals of other organic disease, such as celiac disease. Therefore, the presented numbers of referrals prevented, represent a 'best case scenario' and are applicable as far as IBD is concerned.

A prospective cluster randomised controlled trial is needed to investigate the added value of both POCTs in primary care to effectively refer children for further diagnostic work-up for IBD. Moreover, studies are needed to examine the feasibility and cost-effectiveness of POCT in clinical practice. In our study, the POCTs were not performed at the GP's office, but by experienced laboratory technicians. Therefore, interpretation by an inexperienced GP or medical assistant might be less reliable (e.g. the intensity of the test indicator line can vary). Also, we did not compare the cost and time efficiency of point-of-care testing with sample analyses in a laboratory for ELISA.

CONCLUSION

A diagnostic test strategy in primary care by using a simple POCT calprotectin or lactoferrin has the potential to reduce the need for referral for further diagnostic work-up in specialist care, with a very low risk of missing a child with IBD. Studies are required to investigate, the feasibility, whether POCTs actually will reduce referrals, and cost-effectiveness of the POCTs in children with chronic gastrointestinal symptoms in primary care.

REFERENCES

1. Kim SC, Ferry GD. Inflammatory bowel diseases in pediatric and adolescent patients: Clinical, therapeutic, and psychosocial considerations. *Gastroenterology*. 2004;126(6):1550-1560.
2. de Ridder L, Rings EH, Escher JC, CBO-werkgroep 'IBD bij kinderen'. Guideline 'diagnosis and treatment of inflammatory bowel disease in children'. *Ned Tijdschr Geneeskd*. 2010;154:A1898.
3. Fagerberg UL, Loof L, Myrdal U, Hansson LO, Finkel Y. Colorectal inflammation is well predicted by fecal calprotectin in children with gastrointestinal symptoms. *J Pediatr Gastroenterol Nutr*. 2005;40(4):450-455.
4. Van de Vijver E, Schreuder AB, Cnossen WR, Muller Kobold AC, van Rheenen PF, North Netherlands Pediatric IBD Consortium. Safely ruling out inflammatory bowel disease in children and teenagers without referral for endoscopy. *Arch Dis Child*. 2012;97(12):1014-1018.
5. Henderson P, Anderson NH, Wilson DC. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease: A systematic review and meta-analysis. *Am J Gastroenterol*. 2014;109(5):637-645.
6. van Rheenen PF, Van de Vijver E, Fidler V. Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: Diagnostic meta-analysis. *BMJ*. 2010;341:c3369.
7. Holtman GA, Lisman-van Leeuwen Y, Reitsma JB, Berger MY. Noninvasive tests for inflammatory bowel disease: A meta-analysis. *Pediatrics*. 2016;137(1):1-11.
8. Holtman G.A., Lisman-van Leeuwen Y., Kollen B.J., et al. Diagnostic accuracy of faecal calprotectin for paediatric inflammatory bowel disease in primary care; a prospective cohort study. accepted *Ann Fam Med* 2016. doi: 10.1370/afm.1949.
9. Labaere D, Smismans A, Van Olmen A, et al. Comparison of six different calprotectin assays for the assessment of inflammatory bowel disease. *United European Gastroenterol J*. 2014;2(1):30-37.
10. Walker TR, Land ML, Kartashov A, et al. Fecal lactoferrin is a sensitive and specific marker of disease activity in children and young adults with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2007;44(4):414-422.
11. Zhou X, Xu W, Tang X, et al. Fecal lactoferrin in discriminating inflammatory bowel disease from irritable bowel syndrome: A diagnostic meta-analysis. *BMC gastroenterology*. 2014;14(1):121.
12. Holtman GA, Lisman-van Leeuwen Y, Kollen BJ, et al. Challenges in diagnostic accuracy studies in primary care: The fecal calprotectin example. *BMC Fam Pract*. 2013;14:179-2296-14-179.
13. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol*. 1997;32(9):920-924.
14. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr*. 2014;58(6):795-806.
15. Logan R. Faecal calprotectin for the diagnosis of inflammatory bowel disease. *BMJ*. 2010;341:c3636.
16. Thomson S, Osborn R, Squires D, Reed SJ. International profiles of health care systems 2011: The Commonwealth Fund. http://www.commonwealthfund.org/publications/fund-reports/2011/nov/international-profiles-of-health-care-systems-2011. Published Nov 2011.
17. Naaktgeboren CA, De Groot JA, van Smeden M, Moons KG, Reitsma JB. Evaluating diagnostic accuracy in the face of multiple reference standards. *Ann Intern Med*. 2013;159(3):195-202.
18. Knottnerus JA, Dinant GJ. Medicine based evidence, a prerequisite for evidence based medicine. *BMJ*. 1997;315(7116):1109-1110.

7

19. Joishy M, Davies I, Ahmed M, et al. Fecal calprotectin and lactoferrin as noninvasive markers of pediatric inflammatory bowel disease. *J Pediatr Gastroenterol Nutr*. 2009;48(1):48-54.
20. Otten CM, Kok L, Witteman BJ, et al. Diagnostic performance of rapid tests for detection of fecal calprotectin and lactoferrin and their ability to discriminate inflammatory from irritable bowel syndrome. *Clinical chemistry and laboratory medicine*. 2008;46(9):1275-1280.
21. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: Assessing new tests against existing diagnostic pathways. *BMJ*. 2006;332(7549):1089-1092.
22. Pels LP, Van de Vijver E, Waalkens HJ, et al. Slow hematological recovery in children with IBD-associated anemia in cases of "expectant management". *J Pediatr Gastroenterol Nutr*. 2010;51(6):708-713.
23. Mack DR, Langton C, Markowitz J, et al. Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics*. 2007;119(6):1113-1119.

Appendix 1. The prevalence of positive calprotectin and lactoferrin POCTs by diagnosis

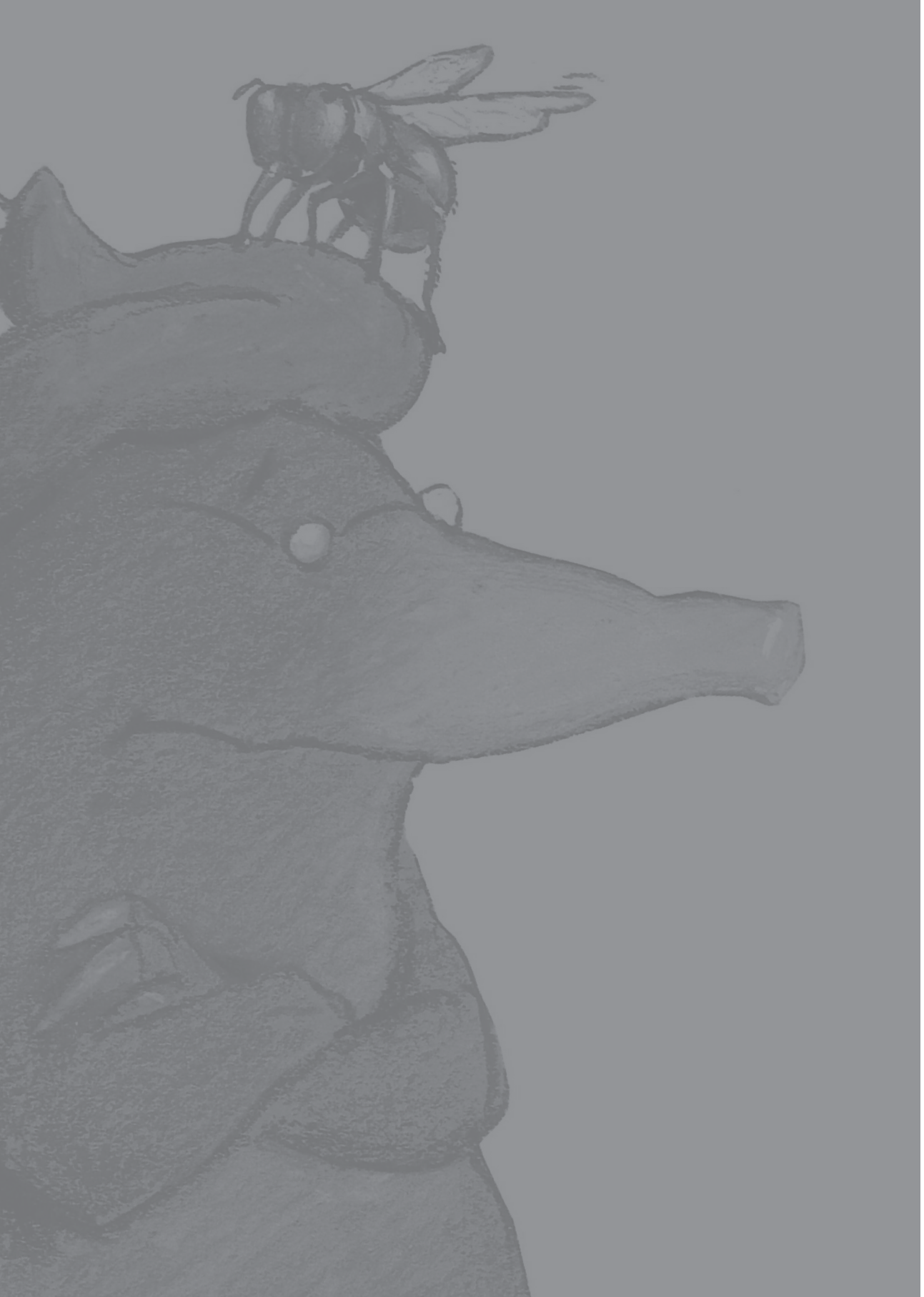
Diagnosis	N (%)	FCal POCT	Flacto POCT
Primary care cohort			
Functional gastrointestinal disorder	108 (95)	4/100	2/100
Gastroenteritis[a]	5 (45)	1	0
Refused endoscopy	1 (1)	-	-
Referred cohort			
IBD			
Crohn's disease	7 (8)	6/6	6/6
Ulcerative colitis	8 (9)	7	7
IBD unclassified	2 (2)	2	2
Non-IBD			
Functional gastrointestinal disorder	66 (73)	4/63	0/63
Gastroenteritis[a]	3 (3)	0	0
Reflux oesophagitis	1 (1)	0	0
Celiac disease	1 (1)	0	0
Solitary rectum ulcer	1 (1)	1	1

Abbreviations: FCal, faecal calprotectin; FLacto, faecal lactoferrin; POCT, point-of-care test; IBD, inflammatory bowel disease.

[a]gastroenteritis by *salmonella enteric* (primary care cohort n = 0; referred cohort n = 2), *Shiga Toxin-Producing Escherichia Coli* (STEC) (n = 1; n = 0), *Giardia lamblia* (n = 4; n = 1).

Note: In one child the diagnosis was unknown, since the child refused endoscopic evaluation at baseline and an indicated 12 months follow-up evaluation by a GP because of presence of chronic gastrointestinal symptoms.

7



CHAPTER 8

SUMMARY AND GENERAL DISCUSSION

The main objective of this thesis was to study the diagnostic strategies for inflammatory bowel disease (IBD) in children with chronic gastrointestinal symptoms, focusing on the value of testing with faecal calprotectin in primary care. This chapter summarises the main results, discusses clinical implications for primary care and methodological implications for diagnostic studies in primary care, and presents suggestions for further research.

SUMMARY OF MAIN FINDINGS

In **chapter 2**, a systematic review and meta-analysis is presented with an overview of the diagnostic accuracy of the signs, symptoms and diagnostic tests for IBD available in all healthcare settings. All 19 included studies were performed in secondary or tertiary care, which resulted in a high disease prevalence, ranging from 19% to 82%. It was shown, that for those children in whom a paediatrician considered endoscopy, symptoms were not sufficiently accurate to distinguish complaints caused by IBD from those related to other causes. Therefore, simple tests that produce minimal adverse effects were recognized as being essential when triaging for endoscopy. With a pooled negative likelihood ratio of 0.01 (0.0–0.1), faecal calprotectin had demonstrable ability in decreasing the probability of IBD. In addition, c-reactive protein and albumin had pooled positive likelihood ratios sufficiently high to indicate an increase of the probability of IBD that may be of clinical importance, with positive likelihood ratios of 5.1 (2.8–9.4) and 8.3 (3.7–18.7), respectively. Thus, we showed that a normal faecal calprotectin could safely rule out IBD, while a positive c-reactive protein or albumin could rule in IBD.

In **chapter 3** we investigated the added diagnostic value of individual blood markers and faecal calprotectin beyond the value of symptoms alone. The results in this section were based on individual patient data in eight studies of referred children. Faecal calprotectin added the most diagnostic value to symptoms when compared against commonly used blood markers. When faecal calprotectin was added to symptoms, the proportion of the total number of patients assigned to the intermediate risk group, with a IBD probability ranging from 35% to 60%, decreased from 55% to 6%. Thus, adding faecal calprotectin to the diagnostic work-up of children with symptoms suggestive of IBD considerably decreased challenging diagnosis.

An important finding of the studies in previous chapters was that none was specifically performed among children in primary care. This is important because the diagnostic accuracies of tests performed in specialist care are not generalizable to primary care. In fact, differences in the patient spectrum and disease severity can affect pre-test probabilities and diagnostic test characteristics, including their sensitivity and specificity values.[1] This highlighted the need for further studies among children in primary care.

In **chapter 4** we described the methodological challenges of performing a diagnostic study in primary care. The DOK study was designed to evaluate the diagnostic accuracy of faecal calprotectin for IBD in children presenting with chronic gastrointestinal symptoms in primary care. Two prospective cohorts of children with chronic gastrointestinal symptoms were included: children presenting in primary care (primary care cohort) and children referred to specialist care (referred cohort). Faecal calprotectin was measured at inclusion and compared by endoscopic assessment or diagnosis at one-year clinical follow-up, which were used as reference standards for IBD diagnosis. Physicians were blinded to the faecal calprotectin results, and calprotectin values above 50 µg/g faeces were considered positive. Two important methodological

challenges were noted: not only does IBD have a low prevalence in primary care, but it is also diagnosed by invasive endoscopy under specialist care. A pragmatic design was therefore used.

Given the low prevalence of paediatric IBD in children with chronic gastrointestinal symptoms presenting in primary care, an ideal cohort study would require long term investigation in a very large population to identify a sufficient number of children with IBD. Given a prevalence of IBD of approximately 1% among children with chronic gastrointestinal symptoms presenting in primary care, we calculated that 7300 symptomatic children would be needed to estimate the sensitivity of faecal calprotectin with adequate precision. In such a case, the financial and logistic difficulties are prohibitive, making such a study infeasible. Therefore, we included a cohort of children in primary care and a cohort of referred children. The prevalence of IBD is known to be around 20% among children referred to specialist care, and this higher prevalence allowed the number of children needed to estimate the sensitivity to be reduced to 90. However, because a higher prevalence of IBD was anticipated (approximately 60%), the precision of the 95% confidence interval for the sensitivity decreased from 5% to 10%.

Children with a low likelihood of IBD were followed for 12 months because it was considered unethical to expose this group to invasive endoscopy. This choice could introduce a differential verification bias, where the two reference standards may have produced differences in classifications of IBD diagnosis that could influence the results for sensitivity and specificity. The verification pattern and the length of the follow-up are important aspects for deciding whether the verification introduced clinical relevant bias. Although the verification pattern was based on the clinical judgements of paediatric gastroenterologist, and we may assume that these assessments were at least somewhat subjective, our results showed that children who underwent endoscopy at baseline were at a higher risk for IBD than children who received follow-up. A 12-month follow-up period was considered appropriate to observe the appearance of new alarm symptoms suggestive for IBD because it is very rare for untreated cases to remain symptom-free beyond that time. Although follow-up was the best achievable option given the reality of clinical care, there remains a small chance that we misclassified cases. Therefore, in chapter 5, we provide insight into pattern and extent of differential verification by comparing the contingency tables by reference standard.

In **chapter 5** we reported the test characteristics of faecal calprotectin for IBD in children presenting with chronic gastrointestinal symptoms in primary care. In the primary care cohort, 24 children had more than one red flag and were referred to a paediatric gastroenterologist. None of the children in the primary care cohort was diagnosed with IBD. Faecal calprotectin was elevated (>50 µg/g faeces) in 15 of 114 children, yielding an overall specificity of 0.87 (0.80–0.92). The referred cohort consisted of 24 children with red flags presenting in primary care, plus 66 children referred to specialist care by a general practitioner or general paediatrician. In these 90 referred children, IBD was confirmed by endoscopy in 17 cases (19%). All children with IBD had elevated faecal calprotectin levels (range 53–2823 µg/g faeces), giving a sensitivity of 0.99 (0.81–1.00) for faecal calprotectin. We concluded that a positive faecal calprotectin result in children with chronic gastrointestinal symptoms presenting in primary care was unlikely to indicate IBD by itself. However, in children in whom a general practitioner considers referral for further diagnostic assessment of IBD, a normal faecal calprotectin result could safely rule out IBD and reduce the number of referrals.

In **chapter 6** we evaluated the added diagnostic value of c-reactive protein and faecal calprotectin beyond evaluation of alarm symptoms, to determine the optimal diagnostic test strategy before referral to specialist care among children with symptoms suggestive of IBD. We compared three test strategies: 1) alarm symptoms alone, based on history and physical examination findings; 2) alarm symptoms plus c-reactive protein; and 3) alarm symptoms plus faecal calprotectin. Of the 90 included children, 17 (19%) had IBD. Adding the faecal calprotectin result to the presence of alarm symptoms increased the area under the curve significantly from 0.80 (0.69–0.90) to 0.97 (0.93–1.00) (P = 0.002). However, adding c-reactive protein to alarm symptoms did not increase the area under the curve significantly. Decision curves confirmed these patterns and showed that the addition of faecal calprotectin to the clinical evaluation of alarm symptoms provided the diagnostic test strategy with the highest net benefit at all reasonable threshold probabilities. We concluded that the optimal strategy, for further stratifying children who have already been identified as at risk for IBD by their general practitioner, was to evaluate the presence of alarm symptoms and faecal calprotectin together.

In **chapter 7** we explored the test characteristics of a point-of-care test measuring both faecal calprotectin and lactoferrin for IBD in children presenting in primary care with chronic gastrointestinal symptoms. In the primary care cohort, calprotectin and lactoferrin had specificities of 0.95 (0.89–0.98) and 0.98 (0.93–0.99), respectively. In the referred cohort, the sensitivities of calprotectin and lactoferrin were both 0.94 (0.72–0.99). Surprisingly, use of the tests together provided no substantial added value over the use of either in isolation. The use of point-of-care calprotectin testing could reduce the referral rate by 76%, while the use of the lactoferrin test did so by 81%, but at the expense of missing one child with IBD (6%). Thus, point-of-care testing has the potential to reduce the need for referral for further diagnostic work-up in specialist care, with a low risk of missing children with IBD.

CLINICAL IMPLICATIONS FOR PRIMARY CARE

It is a diagnostic challenge in primary care to differentiate between functional gastrointestinal disorders, in which diagnostic testing should be minimized, and organic pathology that should not be missed. IBD and coeliac disease are the most important diagnoses a general practitioner needs to eliminate before considering a functional disorder. At the start of this thesis, several guidelines that were designed to assist the general practitioner in this diagnostic dilemma were reviewed. These ranged from general guidance for the management of abdominal pain in children to more specialized guidance for the diagnosis and treatment of children with IBD.[2-4] Although small discrepancies exist between these guidelines, they consistently suggest, that when evaluating the risk of IBD in children with chronic gastrointestinal symptoms, the general practitioner should be guided by the presence or absence of alarm symptoms and the results of blood tests.

The main results of this thesis, that faecal calprotectin had good test characteristics in primary care and added value over alarm symptoms alone, suggest that faecal calprotectin can be used as a triage test to reduce unnecessary referrals among symptomatic children. However, this is contingent on the faecal calprotectin test not being used as a replacement for a thorough history and physical examination. We also found that c-reactive protein had no additional value over alarm symptoms, and that faecal calprotectin showed the

highest discriminative power of all laboratory markers, indicating that faecal calprotectin is presently the best triage test for IBD in primary care. In theory, it is possible that the faecal calprotectin test could replace the need for invasive blood tests in children with suspected IBD. When coeliac disease is suspected, antibody tissue transglutaminase (anti-TTG) and immunoglobulin A tests can also be performed.[5] In this study all anti-TTG tests were normal, including that in the case of a child with coeliac disease and immunoglobulin A deficiency.

Given the very low probability of IBD and the considerable risk of false positives, faecal calprotectin cannot be recommended for use in the screening of all children with chronic gastrointestinal symptoms. The number of false positives (13%) in our study was comparable to the previously reported number of false positives (10%) in children with functional gastrointestinal disorders presenting to specialist care.[6] Referral of children with functional gastrointestinal disorders should be avoided because this approach does not lead to the intended reassurance for either child or parent, and it may in fact lead to a chronic course of abdominal pain.[7,8] In addition, the decision curve analysis showed that testing for faecal calprotectin in children with a threshold probability below 5% barely improved the net benefit compared to the net benefit of referring all children. This highlights the importance of testing only in the context of specific clinical features, and of adopting a sequential strategy to ensure appropriate referral for further diagnostic assessment. Testing with faecal calprotectin can only be recommended in children in whom the general practitioner has identified alarm symptoms; that is, those children in whom they consider referral for further assessment of possible IBD.

To maintain the high negative predictive value, the diagnostic threshold might need to be higher in primary care.[9] Although, in chapter 5, we found that an increase in the threshold from 50 µg/g to 250 µg/g faeces reduced the referral rate by 14% (with a drop from 32% to 18% referred), this threshold also led to false-negative results and missed cases with IBD (with an increase from 0% to 18% cases missed). Therefore, in children for whom the general practitioner considers referral for diagnostic evaluation for IBD, they can reassure those with a calprotectin value below 50 µg/g and refer to specialist care those with a value above 250 µg/g faeces. In children with a value between 50 µg/g and 250 µg/g faeces, the calprotectin value alone gives no direct justification for referral, but this does not preclude the chance that the child may have or may develop IBD or another organic disease.[10] A safety net policy could be applied in this situation, with clear communication given to children and their parents about the situation, including advice on the need to re-contact the general practitioner if specific signs or symptoms develop and reassurance that children with persistent symptoms will be re-tested. Children whose symptoms persist and whose calprotectin values remain high can then be referred during the course of follow-up in primary care. A disadvantage of this strategy is, that while unnecessary referral may be prevented in children with functional gastrointestinal disorders, diagnosis may be delayed in children with IBD. Figure 1 shows a hypothetical diagnostic strategy for IBD in children presenting with chronic gastrointestinal symptoms in primary care.

The faecal calprotectin test is limited by the fact that levels increase after the use of non-steroidal anti-inflammatory drugs and antibiotics.[11] In our study, children using this medication for more than 3 months were excluded, and children using the medication for <3

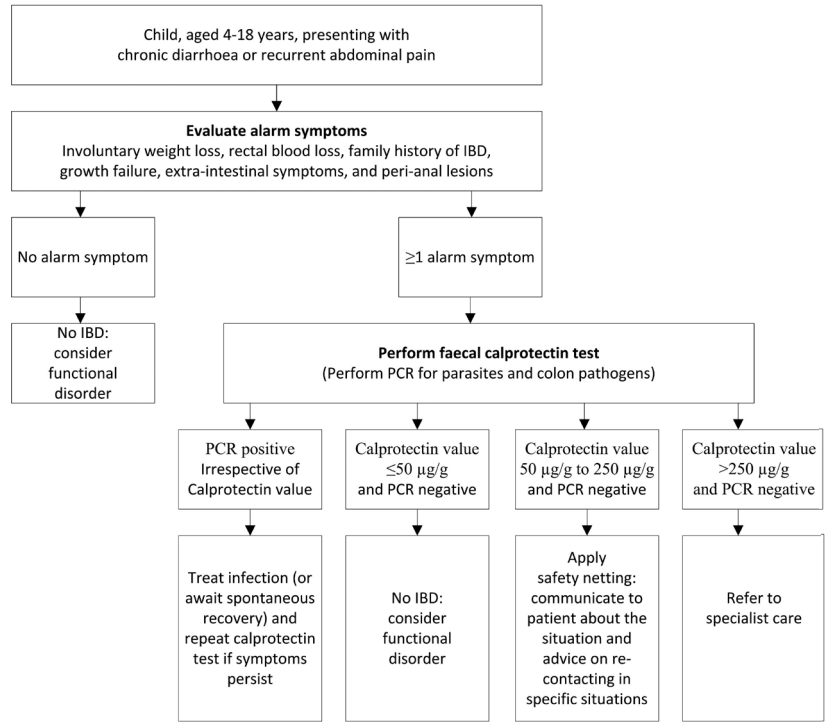


Figure 1. Hypothetical diagnostic strategy for IBD in children presenting with chronic gastrointestinal symptoms in primary care.

Note: Chronic diarrhea is defined as soft to watery stool (score 5, 6, or 7 on the Bristol stool chart) for ≥2 weeks or ≥2 episodes in the past 6 months. Recurrent abdominal pain is defined as ≥2 episodes of abdominal pain or discomfort in the past 6 months.

months had stool collection postponed until the end of treatment. Three children without IBD used antibiotics (n = 2) or NSAIDs (n = 1) for 7 days at baseline and had their stool samples collected at the end of treatment. Two of these three children displayed elevated calprotectin values, which resulted in false-positive results. More studies are required to evaluate the optimal interval needed between the end of treatment and testing with faecal calprotectin.

In a correspondence letter, Poullis et al.[12] suggested that proton-pump inhibitors also affect the faecal calprotectin level. Since we did not include any children using proton-pump inhibitors, we could not evaluate this hypothesis. In total, 43 children reported chronic medication use, typically the oral contraceptive pill or asthma medication (salmeterol/fluticasone), but we did not evaluate whether these affected faecal calprotectin levels because none of the recorded medications was associated with false-positive results in previous studies. Moreover, although the calprotectin level is known to be higher in healthy children younger than four years and in children with bacterial gastroenteritis,[13,14] only one of the eight children with gastroenteritis in the DOK study had an elevated calprotectin value. Nevertheless, it is important that the general practitioner be aware of these factors that

influence calprotectin levels.

There is presently insufficient evidence on either the diagnostic value or the cost-effectiveness of point-of-care testing with faecal calprotectin to recommend it over the enzyme-linked immunosorbent assay (ELISA) test. However, the point-of-care test showed comparable test characteristics as the ELISA test in the same patient population, though more studies are needed to confirm this result. The characteristics of the ELISA test are reported in chapter 5, while those of the point-of-care test are reported in chapter 6. An important benefit of the point-of-care test is that stool samples can be brought from home and tested by the general practitioner in clinic, with results available within 15 minutes. This can expedite the diagnostic trajectory for children with chronic gastrointestinal symptoms. However, the labour and equipment costs need to be compared between the point-of-care test and the ELISA test. A pitfall of point-of-care testing is that general practitioner may come to rely on this simple test rather than evaluating alarm symptoms, which might lead to the unintended consequence of an increased number of false-positive results and, therefore, an increased number of unnecessary referrals.

METHODOLOGICAL IMPLICATIONS FOR DIAGNOSTIC STUDIES IN PRIMARY CARE

Patient selection

The characteristics of a test are evaluated by comparing the results of the test to be evaluated with the results of a reference standard in the same patients. Ideally, this should involve a cohort of consecutive patients with clinical suspicion of having the condition of interest, and in which both the test under evaluation and the reference standard are performed in all patients. Studies among patients with low prior disease probabilities have important methodological challenges, specifically that the inclusion of a proper amount of patients with the disease requires the inclusion of a large number of symptomatic patients. It is important to identify methodological solutions that reduce the sample size and costs of the investigation, deal with ethical challenges, yet yield unbiased estimates of test accuracy.

So, what are the advantages and disadvantages of the pragmatic design used in the DOK study of a primary care and referred cohort? An obvious advantage is that we could compare the specificity in both cohorts. Moreover, by performing subgroup analysis in children who were referred by their general practitioner and paediatrician, we could gain insight into the influence of prevalence on the sensitivity and specificity of faecal calprotectin. We showed that spectrum bias did affect the specificity of faecal calprotectin, but that it did not substantially affect the sensitivity. The tendency of the specificity to be higher at a lower disease prevalence, and for a sensitivity that did not change by disease prevalence, was comparable with the result of a meta-analysis investigating the association between test sensitivity and specificity with disease prevalence.[1] This implies that differences in prevalence particularly reflect differences in the spectrum of patients without disease. Two of the twenty-three included meta-analyses showed significant associations between prevalence and sensitivity, with sensitivity being higher among patient populations where the disease prevalence was higher. Therefore, the authors of the meta-analysis concluded that clinicians can use the prevalence as a guide to select studies that closely match their setting.

A disadvantage of the patient selection method was that the inclusion of children with IBD was lower than expected. If we had included patients with IBD at a prevalence of 20%, the precision of the 95% confidence intervals would have decreased from 5% to ±10%. Because the prevalence of IBD in referred children was 19%, the sensitivity had wide confidence intervals ranging from 0.81 to 1.00. The implications of a sensitivity of 0.81 might be of concern to patients and clinicians. Further research in primary care might have an important impact on our confidence in the effect estimate, causing it to change.

Although diagnostic studies concerning rare diseases in primary care are difficult, other solutions are available to deal with the methodological challenge of patient selection. This includes case-control studies with reversed-flow or two-gate design with representative sampling (i.e., a nested case-control design). Using registration data for routine care might be a solution to help determine the diagnostic accuracy of faecal calprotectin in symptomatic children presenting in primary care. But, such a study would only be feasible if calprotectin was used in routine care. Also, registration data often does not reliably include the details about symptoms and tests necessary for high-quality diagnostic research. However, solutions for the missing data and other methodological challenges have been proposed in diagnostic test research.[15] A nested case-control design, using registration data, might be an underused solution for determining the diagnostic accuracy of studies in patients with a low prior disease probability in primary care. In such a nested case-control design, the cases and controls are extracted from a source population with a known sample size. Compared with the traditional case-control design, it has the advantage of being suitable for calculating the absolute disease risk next to predictive values and odds ratios; thus, it might be useful in diagnostic accuracy studies among patients with low probability of having the disease.[16] Biases associated with these solutions have been extensively reported,[17-19] but an overview describing methodological challenges linked to practical recommendations, and how to deal with them, is lacking for diagnostic studies among patients with low prior disease probabilities. Recommendations for researchers on design, analysis, reporting and interpreting of results are needed.

Multiple reference standards

Verification is a critical step in any study of diagnostic accuracy. Ideally, the preferred reference standard should be performed in all symptomatic patients. However, diagnosis is not always possible with the preferred reference standard for practical or ethical reasons. In these situations, an alternative reference standard can be used. Although this method of differential verification can be a useful solution, bias can be introduced if the inferior reference standard incorrectly classifies patients as diseased or non-diseased; the resulting diagnostic accuracy would then be systematically different from that in an ideal study using only the preferred reference standard. This systematic deviation introduced by the use of two reference standards is called differential verification bias.[20]

Naaktgeboren et al. 2013 provided practical recommendations on how to report results when differential verification occurs.[21] Our presentation of results in chapter 5 was based on these recommendations. A flow diagram, such as that presented in chapter 5, gives an example of how to report the verification process and contingency tables for the index test by each reference standard. Moreover, they provide additional questions related to the QUADAS-2

(the revised version of the tool for the quality assessment of diagnostic accuracy studies) for assessing the risk of clinically relevant differential verification bias. The four questions are:

- Was the choice of reference standard completely dependent on the results of the index test? (if so, the predictive values are clinically interpretable.)
- If the answer to the first question is no, how accurate is the inferior reference standard? (The higher the accuracy of the inferior reference standard, the lower the risk of bias.)
- What percentage of the participants were diagnosed by use of the inferior reference standard? (If a negligible percentage of participants received an inferior standard, the risk of bias is low. Several factors must be taken into account when determining whether the percentage is negligible.)
- If follow-up is used as the inferior reference standard, does it identify almost all hidden cases present at the time of the index test, but very few new cases that developed afterward? Does follow-up detect the same type of cases as the preferred reference standard? (If the answer to both questions is yes, the risk of bias is low.)

When assessing these questions for the DOK study, the risk of bias was estimated as low to moderate. The risk of bias with 12 months' follow-up was considered low because IBD is not a self-limiting disease and typically presents with alarm symptoms within a few months of clinical suspicion. Therefore, the probability of IBD would be extremely low for a child without either alarm symptoms or an indication for endoscopy during 12 months' follow-up. Because the choice of the reference standard was not dependent on the index test, and because a large percentage of the participants were diagnosed with only the 12-month follow-up, the risk of bias was considered moderate.

In chapter 6 we evaluated the verification pattern to show that children who underwent endoscopy at baseline were at higher risk for IBD than children who were followed for 12 months. Therefore, we concluded that the results would probably have been biased if the test characteristics were only evaluated among children who received endoscopy.[22] A disadvantage of this approach is that we could not use a Bayesian correction method to adjust for the imperfection of using the 12-months' follow-up, because the verification pattern of children was not entirely clear between those who received endoscopy and those who received 12 months' follow-up. When designing future studies to use the Bayesian correction method, it will be essential to determine the verification pattern before starting the study. Moreover, information on the diagnostic accuracy of the reference standard will be needed.[20]

SUGGESTIONS FOR FURTHER RESEARCH

In this thesis we focused on the patient perspective by evaluating the number of false negatives (missed diagnosis) or false positives (referrals).[23] The results suggest that, in theory, implementation of faecal calprotectin testing when alarm symptoms are present can reduce the number of referrals without missing cases of IBD. A reduction in referrals for secondary care and expensive and invasive diagnostic procedures, combined with a more rapid diagnosis and management when IBD is present, would greatly benefit children, physicians, and policy makers when faced with chronic gastrointestinal symptoms. However, it is important to note that other factors might influence the clinical decisions of a general practitioner in daily practice; for example, parents may insist on the need for further investigation.[24] Moreover, we did not evaluate the population perspective, so we did not assess how the test would affect health care budgets. Both these factors should be addressed in future research. Furthermore, it would be interesting to see research evaluating the diagnostic value of the Rome III criteria in combination with the exclusion of alarm symptoms and normal faecal calprotectin when evaluating functional gastrointestinal disorders, such as irritable bowel syndrome.

An impact or implementation study is needed to determine whether using faecal calprotectin in the presence of alarm symptoms actually improves decision making and cost-effectiveness in daily practice in primary care. This strategy could be evaluated by a stepped wedge cluster randomized trial or a before–after trial in which the diagnostic strategy is evaluated before and after implementing the strategy in a large number of general practices. Because the low prevalence of IBD could make a randomized trial difficult to achieve, the use of before and after registration data might be a solution. Other relevant solutions have been proposed for methodological problems associated with the use of routine care databases in observational studies on treatment interventions (such as missing data and confounding by indication), and these may also apply to diagnostic test research.[15,25,26]

Outcomes could be set as the number of appropriate or inappropriate referrals, or the impact of those referrals on quality of life. However, defining what constitutes appropriate or inappropriate referrals would be difficult; for example, although we focused on IBD in this thesis, referrals for coeliac disease, lactose intolerance, and other organic diseases might be considered appropriate, as might the referral of children with functional symptoms when management is proving difficult. The latter case also emphasizes that referral can serve aims other than for further diagnostic work-up in children with suspected organic disease.

Once impact studies have confirmed whether faecal calprotectin is a useful test in primary care, it will be possible to include it in primary care guidelines, which should use the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach to assessing the quality of evidence and strength of recommendations.[23] The GRADE approach allows accurate research and accurate debate concerning judgements about diagnostic tests, and the evidence profiles provide simple transparent summaries that can be used by clinicians.[27]

OVERALL CONCLUSION

In this thesis, we have shown that faecal calprotectin is a useful test for ruling out IBD in children with chronic gastrointestinal symptoms in whom the general practitioner considers referral for further diagnostic work-up, e.g. in the presence of alarm symptoms. Our data indicated that optimal outcomes in the management of children presenting with chronic gastrointestinal symptoms in primary care were achieved by adopting a sequential diagnostic strategy. Specifically, we demonstrated that a strategy of performing faecal calprotectin testing when alarm symptoms had been identified was associated with a reduction in unnecessary referrals. However, an impact study is now needed to determine whether this approach might actually improve both the diagnostic decision making of general practitioners in daily practice and the cost-effectiveness of diagnostic assessment in primary care.

REFERENCES

1. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ*. 2013;185(11):E537-44.
2. Levine A, Koletzko S, Turner D, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr*. 2014;58(6):795-806.
3. de Ridder L, Rings EH, Escher JC, CBO-werkgroep 'IBD bij kinderen'. Guideline 'diagnosis and treatment of inflammatory bowel disease in children'. *Ned Tijdschr Geneeskd*. 2010;154:A1898.
4. Gieteling M, Van Dijk P, De Jonge A, Albeda F, Berger M, Burgers J. NHG-standaard buikpijn bij kinderen. *Huisarts Wet*. 2012;55:404-409.
5. Hill ID, Dirks MH, Liptak GS, et al. Guideline for the diagnosis and treatment of celiac disease in children: Recommendations of the north american society for pediatric gastroenterology, hepatology and nutrition. *J Pediatr Gastroenterol Nutr*. 2005;40(1):1-19.
6. Flagstad G, Helgeland H, Markestad T. Faecal calprotectin concentrations in children with functional gastrointestinal disorders diagnosed according to the pediatric rome III criteria. *Acta Paediatr*. 2010;99(5):734-737.
7. Gieteling MJ, Bierma-Zeinstra SM, Passchier J, Berger MY. Prognosis of chronic or recurrent abdominal pain in children. *J Pediatr Gastroenterol Nutr*. 2008;47(3):316-326.
8. van Tilburg MA, Venepalli N, Ulshen M, Freeman KL, Levy R, Whitehead WE. Parents' worries about recurrent abdominal pain in children. *Gastroenterology Nursing*. 2006;29(1):50-55.
9. Logan R. Faecal calprotectin for the diagnosis of inflammatory bowel disease. *BMJ*. 2010;341:c3636.
10. Buntinx F, Mant D, Van den Bruel A, Donner-Banzhof N, Dinant GJ. Dealing with low-incidence serious diseases in general practice. *Br J Gen Pract*. 2011;61(582):43-46.
11. Gisbert J, McNicholl A. Questions and answers on the role of faecal calprotectin as a biological marker in inflammatory bowel disease. *Digestive and Liver Disease*. 2009;41(1):56-66.
12. Poullis A, Foster R, Mendall MA. Proton pump inhibitors are associated with elevation of faecal calprotectin and may affect specificity. *Eur J Gastroenterol Hepatol*. 2003;15(5):573&hyhen.
13. Oord T, Hornung N. Fecal calprotectin in healthy children. *Scand J Clin Lab Invest*. 2014;74(3):254-258.
14. Duman M, Gencpinar P, Biçmen M, et al. Fecal calprotectin: Can be used to distinguish between bacterial and viral gastroenteritis in children? *Am J Emerg Med*. 2015;33(10):1436-1439.
15. Oostenbrink R, Moons KG, Bleeker SE, Moll HA, Grobbee DE. Diagnostic research on routine care data: Prospects and problems. *J Clin Epidemiol*. 2003;56(6):501-506.
16. Biesheuvel CJ, Vergouwe Y, Oudega R, Hoes AW, Grobbee DE, Moons KG. Advantages of the nested case-control design in diagnostic research. *BMC Med Res Methodol*. 2008;8:48-2288-8-48.
17. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174(4):469-476.
18. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem*. 2005;51(8):1335-1341.
19. Obuchowski NA, Zhou XH. Prospective studies of diagnostic test accuracy when disease prevalence is low. *Biostatistics*. 2002;3(4):477-492.
20. de Groot JA, Dendukuri N, Janssen KJ, Reitsma JB, Bossuyt PM, Moons KG. Adjusting for differential-verification bias in diagnostic-accuracy studies: A bayesian approach. *Epidemiology*. 2011;22(2):234-241.
21. Naaktgeboren CA, De Groot JA, van Smeden M, Moons KG, Reitsma JB. Evaluating diagnostic accuracy in the face of multiple reference standards. *Ann Intern Med*. 2013;159(3):195-202.
22. Naaktgeboren CA, de Groot JA, Rutjes AW, Bossuyt PM, Reitsma JB, Moons KG. Anticipating missing reference standard data when planning diagnostic accuracy studies. *BMJ*. 2016;352:i402.
23. Schunemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ*. 2008;336(7653):1106-1110.
24. Heida A, Holtman GA, Lisman-van Leeuwen Y, Berger MY, van Rheenen PF. Avoid endoscopy in children with suspected inflammatory bowel disease who have normal calprotectin levels. *J Pediatr Gastroenterol Nutr*. 2016;62(1):47-49.
25. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev*. 2014;4.
26. Cousens S, Hargreaves J, Bonell C, et al. Alternatives to randomisation in the evaluation of public-health interventions: Statistical analysis and causal inference. *J Epidemiol Community Health*. 2011;65(7):576-581.
27. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. GRADE: What is "quality of evidence" and why is it important to clinicians. *Chinese Journal of Evidence-Based Medicine*. 2009;2:009.

8

NEDERLANDSE SAMENVATTING

Het doel van dit proefschrift is het onderzoeken van verschillende diagnostische strategieën voor kinderen met een verdenking van een chronische inflammatoire darmziekte (IBD). De focus ligt daarbij op het onderzoeken van de (meer) waarde van het meten van het ontstekingseiwit 'calprotectine' in de ontlasting bij kinderen die op het spreekuur van de huisarts worden gezien met langdurige buikpijn en/of diarree. In dit hoofdstuk worden in begrijpelijk Nederlands de achtergronden van het diagnostisch dilemma, het literatuuronderzoek naar de diagnostische testen voor IBD, de opzet van de DOK studie, de belangrijkste bevindingen van de DOK studie en de algemene conclusies en aanbevelingen besproken.

ACHTERGROND

Op het spreekuur van de huisarts komen regelmatig kinderen met langdurige buikpijn en/of diarree. De oorzaak kan functioneel of organisch zijn. We spreken van functionele klachten als er geen onderliggende oorzaak wordt gevonden. Kinderen met functionele buikklachten worden vaak gezien door de huisarts. Het beleid bij deze klachten, vervolgen, psycho-educatie en gepaste triage, hoort bij uitstek tot het domein van de huisarts. Het snel herkennen van functionele klachten en het leren omgaan met deze klachten kan ervoor zorgen dat de kwaliteit van leven van de kinderen en de prognose van de klachten verbeteren. Aan de andere kant is het belangrijk om een organische oorzaak, zoals bijvoorbeeld IBD, niet te missen. Kinderen met IBD hebben een chronische ontsteking van het maag darm stelsel. De twee hoofdtypen zijn colitis ulcerosa en de ziekte van Crohn. IBD komt maar heel weinig voor, maar het tijdig herkennen en behandelen van deze aandoening kan ernstige gevolgen, zoals groeivertraging en bloedarmoede, voorkomen. Het bepalen of iemand IBD heeft wordt gedaan met een darmonderzoek (endoscopie) en dit wordt uitgevoerd door een kinderarts Maag Darm Lever (MDL). Bij kinderen vindt dit onderzoek plaats onder volledige anesthesie (narcose).

Voor de huisarts is het belangrijk om alléén de kinderen met een grote kans op IBD te verwijzen naar een specialist, zodat enerzijds vertraging in diagnostiek voorkomen wordt, en anderzijds onnodige endoscopieën bij kinderen zonder IBD voorkomen kunnen worden. Het onderscheid tussen functionele en organische aandoeningen is niet altijd te maken aan de hand van het symptomenpatroon. De huisarts kan tot op heden gebruik maken van bloedmarkers om de waarschijnlijkheid van IBD beter in te schatten. Deze testen zijn echter niet specifiek voor het vaststellen van een ontsteking in de darmen en maken daarom niet altijd goed onderscheid tussen kinderen mét en zonder IBD. De test 'fecaal calprotectine' is een eenvoudige en niet invasieve test, welke wel de potentie heeft om een goed onderscheid te maken tussen kinderen mét en zonder IBD. De test fecaal calprotectine meet de concentratie calprotectine in de ontlasting. Calprotectine is een eiwit dat bij een ontsteking vrijkomt. Bij een verhoogde waarde van calprotectine in de ontlasting heeft het kind waarschijnlijk een hogere kans op IBD. Deze test wordt al gebruikt door kinderartsen in het ziekenhuis, maar nog niet in de huisartsenpraktijk.

DIAGNOSTISCHE TESTEN

Naast bloedtesten en fecaal calprotectine zijn er nog andere testen, zoals andere fecestesten, urinetesten en echografie die de huisarts zou kunnen helpen bij het bovengeschetste diagnostische dilemma.

Hoofdstuk 2 presenteert de resultaten van een systematisch literatuuronderzoek en

meta-analyse van studies die de diagnostische waarde van symptomen en eenvoudige testen voor IBD onderzochten. We vonden in de literatuur 19 publicaties van voldoende kwaliteit om de resultaten te kunnen bundelen. Deze onderzoeken zijn allen uitgevoerd in het ziekenhuis, dat wil zeggen, in geselecteerde kinderen met een hoge vooraf kans op IBD. Dit zagen we terug in de hoge prevalentie van IBD in de gevonden onderzoeken, IBD kwam bij 19% tot 82% van de totaal onderzochte patiënten voor. Voor deze geselecteerde populatie bleek op basis van symptomen geen onderscheid gemaakt te kunnen worden tussen kinderen mét en zonder IBD. Daarom zijn aanvullende testen belangrijk voor het selecteren van patiënten die daadwerkelijk een endoscopie nodig hebben. Op basis van 10 gepubliceerde studies bleek dat een normale fecaal calprotectine waarde de kans op IBD kan verkleinen met een gepoolde negatieve likelihood ratio van 0.01 (95% BI 0.0-0.01). De bloedtesten C-reactief proteïne (9 studies) en albumine (6 studies) hadden een positieve likelihood ratio boven de 5, wat betekent dat deze twee testen gebruikt kunnen worden om de kans op IBD te verhogen. Dus, een normale fecaal calprotectine waarde kan IBD veilig uitsluiten, terwijl een positieve C-reactieve proteïne of albumine waarde IBD kan insluiten.

In **hoofdstuk 3** wordt de toegevoegde waarde van bloedtesten en fecaal calprotectine bovenop het uitvragen van symptomen en bevindingen uit lichamelijk onderzoek beschreven. Hiervoor zijn individuele patiëntengegevens gebruikt van acht gepubliceerde studies die zijn uitgevoerd in naar het ziekenhuis verwezen kinderen. Als toevoeging aan symptomen geeft fecaal calprotectine in vergelijking met andere diagnostische bloedtesten de meeste informatie. Als we in een hypothetisch cohort van 100 kinderen fecaal calprotectine toevoegen aan symptomen wordt de groep kinderen waarvoor onzekerheid bestaat over de diagnose IBD kleiner, namelijk van 55 naar 6 kinderen. Dus het testen van fecaal calprotectine op een groep verwezen kinderen zou het aantal endoscopieën aanzienlijk kunnen verminderen.

Opmerkelijk is dat geen van de studies is uitgevoerd in de huisartsenpraktijk. Resultaten uit onderzoeken met kinderen uit het ziekenhuis kunnen niet worden geëxtrapoleerd naar de huisartsenpopulatie, door een verschil in ernst van de aandoening en een verschil in case-mix van aandoeningen. Studies in de huisartsenpraktijk zijn nodig om te onderzoeken wat de waarde van de symptomen en eenvoudige testen zijn in de huisartsenpraktijk en wat de (meer) waarde is van fecaal calprotectine.

DE DOK STUDIE

Voordat de fecaal calprotectine test kan worden aanbevolen voor gebruik in de huisartsenpraktijk bij kinderen met langdurige buikpijn en/of diarree moet er onderzoek worden gedaan naar de diagnostische waarde van fecaal calprotectine in deze populatie. Dit is waarom de DOK (*Darm Onderzoek bij Kinderen*) studie is opgezet. Het doel van dit prospectief cohort onderzoek is om de diagnostische waarde van fecaal calprotectine voor IBD te onderzoeken bij kinderen van 4-18 jaar die zich in de huisartsenpraktijk presenteren met chronische buikpijn en/of diarree. Het onderzoek zal een antwoord geven op de vragen “Hoe groot is de kans op een fout-positieve testuitslag bij kinderen met chronische buikklachten in de huisartspraktijk?” en “Hoe groot is de kans op een fout-negatieve testuitslag bij kinderen met chronische buikklachten die de huisarts wil verwijzen, bijvoorbeeld kinderen met alarm symptomen?” Daarnaast zal geëvalueerd worden wat de optimale diagnostische test strategie

is om kinderen met chronische buikklachten te verwijzen.

Voor de DOK studie werden twee cohorten van kinderen geïncludeerd: een ‘huisartsencohort’ en een ‘verwezen cohort’. In het huisartsencohort werden kinderen geïncludeerd die zich presenteren bij de huisarts met chronische buikpijn en/of diarree. Omdat IBD zeldzaam is in de huisartsenpraktijk en om toch genoeg kinderen mét deze aandoening te includeren, is er voor een pragmatische aanpak gekozen. In het verwezen cohort werden daarom kinderen geïncludeerd die naar de kinder-MDL-arts waren verwezen vanuit het huisartscohort, omdat ze één of meer van de zes vooraf gedefinieerde alarmsymptomen hadden of een afwijkend resultaat op één of meer van de vier bloedtesten. Daarnaast werden kinderen geïncludeerd die door andere huisartsen en kinderartsen waren verwezen naar het ziekenhuis voor verder onderzoek van hun chronische darmklachten.

Aan het begin van de studie stuurden alle kinderen een potje met ontlasting op naar het laboratorium, waar het werd opgeslagen bij –80°C. Aan het einde van de studie werd de concentratie fecaal calprotectine bepaald met een laboratorium test en een sneltest. Een sneltest, ook wel point-of-care test genoemd, is een test die meteen uitgevoerd kan worden in de huisartsenpraktijk en het resultaat is bekend binnen 15 minuten. Een waarde boven 50 µg/g feces is positief. Alle artsen, kinderen, ouders en onderzoekers waren geblindeerd voor de calprotectine uitslag.

Als onderdeel van de DOK-studie volgden alle deelnemers een standaard diagnostisch traject om IBD uit te sluiten of aan te tonen. Hierbij werd er gekeken naar zes anamnestische alarmsymptomen: gewichtsverlies, groeivertraging, rectaal bloedverlies, positieve familieanamnese voor IBD, extra-intestinale symptomen en perianale laesies. Ook werd er bloed geprikt voor het bepalen van C-reactief proteïne, bezinking, trombocyten en hemoglobine. De CBO richtlijn “diagnostiek en behandeling van chronische inflammatoire darmziekte bij kinderen” heeft als richtlijn gediend voor het diagnosetraject. Kinderen met een hoog risico op IBD kregen binnen dit traject een endoscopie. Dit is de gouden standaard voor het stellen van de diagnose IBD. Kinderen met een laag risico werden een jaar gevolgd met behulp van vragenlijsten. Na 12 maanden werden alle kinderen die geen endoscopie hadden ondergaan of bij wie bij endoscopie geen IBD was gevonden, nog een keer geëvalueerd door de arts en zo nodig verwezen voor een endoscopie. Als het kind geen klachten meer had die passen bij IBD, werd aangenomen dat het kind de aandoening niet had.

BELANGRIJKSTE BEVINDINGEN

In **hoofdstuk 4** van dit proefschrift wordt de opzet van de DOK studie in detail beschreven. Daarnaast wordt de mogelijke vertekening door twee uitdagende methodologische keuzes in de studie besproken. Om IBD te diagnosticeren wordt een endoscopie gebruikt, maar omdat het niet ethisch is om kinderen met een lage kans op IBD een endoscopie aan te doen worden deze kinderen een jaar gevolgd. Hiermee is er een kans dat er kinderen met IBD onterecht niet worden herkend. Er worden hier dus twee referentiestandaarden gebruikt, wat vertekening van de resultaten kan veroorzaken, dit wordt ook wel *differential verification bias* genoemd. Daarnaast worden in het verwezen cohort twee groepen kinderen geïncludeerd: kinderen die verwezen werden vanuit het huisartsencohort en kinderen die geïncludeerd werden in het ziekenhuis die waren verwezen door andere huisartsen of kinderartsen. Als deze kinderen

hetzelfde zouden zijn, zijn de resultaten generaliseerbaar naar kinderen met een verhoogd risico op IBD in de huisartsenpraktijk. Wanneer deze kinderen verschillen, kan dit *spectrum bias* veroorzaken, waardoor de generaliseerbaarheid van de resultaten moeilijker wordt.

Hoofdstuk 5 presenteert de diagnostische waarde van fecaal calprotectine voor IBD bij kinderen met chronische gastro-intestinale symptomen in de huisartsenpraktijk. In het huisartsencohort hadden 15 van de 114 kinderen fout-positieve uitslagen en dit geeft een specificiteit van 0.87 (95% BI 0.80-0.92). In het verwezen cohort van 90 kinderen waren 17 (19%) kinderen gediagnostiseerd met IBD, waarvan geen enkel kind een fout-negatieve uitslag had. De sensitiviteit was 0.99 (95% BI 0.81-1.00). Deze resultaten wijzen erop dat in kinderen die zich presenteren met chronische gastro-intestinale klachten in de huisartsenpraktijk een positieve fecaal calprotectine waarschijnlijk niet indicatief is voor IBD. Echter, een normale fecaal calprotectine uitslag kan IBD veilig uitsluiten bij kinderen die de huisarts zou willen verwijzen voor verder diagnostische onderzoek voor IBD.

De kans op differential verification bias was klein, omdat er een zeer kleine kans is dat er kinderen met IBD gemist zijn. Het is namelijk zeer onwaarschijnlijk dat een kind langer dan een jaar onherkenbaar rondloopt met IBD. Daarnaast zijn er geen kinderen aan het einde van de 12 maanden follow-up gediagnostiseerd met IBD. Omdat de kinderen die verwezen waren door de kinderarts zieker waren dan kinderen die verwezen waren door de huisarts ontstaat er spectrum bias. De resultaten lieten zien dat spectrum bias nauwelijks de sensitiviteit beïnvloede, maar de specificiteit wel. Dat de specificiteit lager is in meer geselecteerde patiënten en dat de sensitiviteit nauwelijks verandert blijkt ook uit een meta-analyse die de associatie tussen sensitiviteit en specificiteit met ziekte prevalentie onderzoekt.

In **hoofdstuk 6** wordt de toegevoegde waarde van het testen op C-reactief proteïne en fecaal calprotectine op anamnestische alarmsymptomen bestudeerd om een optimale diagnostische strategie voor verwijzing naar de specialist te bepalen in kinderen met een verdenking van IBD. We vergeleken drie diagnostische strategieën: 1) alarm symptomen; 2) alarm symptomen en C-reactief proteïne; 3) alarm symptomen en fecaal calprotectine. Fecaal calprotectine gaf een significante toevoeging op anamnestische alarmsymptomen bij het onderscheiden van kinderen mét en zonder IBD, terwijl het betrekken van C-reactief proteïne geen significante toegevoegde waarde had. De besliscurve van de diagnostische strategieën geeft hetzelfde resultaat en laat zien dat de combinatie van alarm symptomen en fecaal calprotectine het hoogste netto-nut over een bereik van klinisch relevante afkappunten heeft. Het netto-nut geeft de balans weer tussen de voordelen van een terechte diagnose en nadelen van een onterechte diagnose. De afkappunten geven weer hoeveel kinderen de huisarts of kinderarts zou willen verwijzen om één kind met IBD te diagnosticeren. De resultaten geven aan dat fecaal calprotectine in combinatie met alarmsymptomen de optimale diagnostische teststrategie vormen om te bepalen welke kinderen met een verdenking van IBD verwezen moeten worden voor vervolgonderzoek.

Hoofdstuk 7 beschrijft de diagnostische waarde van de sneltest fecaal calprotectine en lactoferrine voor IBD bij kinderen met chronische gastro-intestinale symptomen in de huisartsenpraktijk. Lactoferrine is net als calprotectine een ontstekingsremmend eiwit dat in de ontlasting gemeten kan worden. Fecaal lactoferrine kan mogelijk iets toevoegen op fecaal calprotectine, omdat het een hoge specificiteit laat zien. Dit bleek niet het geval,

omdat in beide cohorten 95% of meer van de kinderen hetzelfde testresultaat hadden voor beide testen. De resultaten laten zien dat beide sneltesten een hoge specificiteit hadden in het huisartsencohort. In het verwezen cohort, hadden beide sneltesten een hoge sensitiviteit en de voorspellende waarde dat er geen IBD in het spel was, bleek ook hoog. De sneltest calprotectine heeft de potentie om het aantal verwijzingen met 76% te verminderen en lactoferrine heeft de potentie om het aantal verwijzingen met 81% te verminderen, terwijl er één kind met IBD werd gemist (6%). Beide sneltesten kunnen verwijzingen voorkomen, met daarbij een laag risico op het missen van een kind met IBD.

ALGEMENE CONCLUSIES EN AANBEVELINGEN

De resultaten van dit proefschrift betekenen dat huisartsen fecaal calprotectine zouden kunnen gebruiken bij kinderen met langdurige buikpijn en/of diarree die ze willen verwijzen naar de kinderarts omdat ze aan IBD denken, bijvoorbeeld omdat het kind ook de bijbehorende alarmsymptomen heeft. In deze kinderen kan een normaal testresultaat op fecaal calprotectine IBD veilig uitsluiten en daarmee verwijzingen voorkomen. Omdat bloedtesten een slechter onderscheidend vermogen hebben dan fecaal calprotectine en het meten van C-reactief proteïne geen toegevoegde waarde laat zien bovenop alarmsymptomen, beïnvloeden bloedtesten de beslissing om een kind wel of niet te verwijzen niet. De sneltest fecaal calprotectine laat een goede diagnostische waarde zien en zou als een betrouwbaar alternatief voor de laboratorium test gebruikt kunnen worden in de huisartsenpraktijk. In vervolg onderzoek is een diagnostische impact studie nodig om te bepalen of het gebruik van fecaal calprotectine in combinatie met anamnestische alarmsymptomen in de dagelijkse praktijk de besluitvorming van huisartsen en de kosteneffectiviteit van de diagnostiek daadwerkelijk verbetert.

Dit proefschrift heeft naast klinische inzichten ook methodologische inzichten opgeleverd. Diagnostisch onderzoek heeft veel methodologische problemen als de kans op aanwezigheid van ziekte klein is en er een invasieve referentie standaard nodig is om de diagnose te stellen. De oplossingen die in dit proefschrift zijn aangedragen zijn voorbeelden van het omgaan met deze methodologische problemen. Methodologische uitdagingen kunnen, wanneer er verkeerd mee om wordt gegaan, vertekende resultaten veroorzaken. Het is erg belangrijk om hiervan bewust te zijn, transparant te zijn in de keuzes die gemaakt worden en deze duidelijk op te schrijven in wetenschappelijke publicaties. Er zijn ook andere mogelijke oplossingen om diagnostisch onderzoek uit te voeren bij patiënten met een lage kans op de ziekte. Deze situatie komt veel voor in de huisartsenpraktijk. Een overzicht van de methodologische uitdagingen gelinkt met praktische aanbevelingen voor onderzoekers is nodig. Dit zal uiteindelijk de kwaliteit van diagnostisch onderzoek in de huisartsenpraktijk, maar ook het diagnostisch proces in de praktijk verbeteren.

DANKWOORD

Dit proefschrift zou er niet geweest zijn zonder de steun en inzet van vele mensen. Daarom wil ik graag iedereen bedanken die heeft meegeholpen aan de totstandkoming van dit proefschrift.

Allereerst wil ik alle **deelnemende kinderen** en hun ouders bedanken voor hun inzet en voor het invullen van vragenlijsten. Ook alle **deelnemende huisartsen** wil ik hartelijk bedanken voor het aanmelden van patiënten en het uitvoeren van het lichamelijk onderzoek. In het bijzonder bedank ik hierbij **Carolien van Leeuwen** voor het organiseren van een nascholingsavond bij haar in de praktijk, grote aantal inclusies van patiënten en een mooie klinische les in het tijdschrift 'huisarts en wetenschap'. De afdelingen kindergeneeskunde (Maag Darm Lever) van het Universitair Medisch Centrum Groningen, Amsterdam Medisch Centrum, Erasmus Medisch Centrum in Rotterdam, Isala Klinieken in Zwolle, Scheper Ziekenhuis in Emmen, Medisch Centrum Leeuwarden, en Wilhelmina Ziekenhuis Assen, heel erg bedankt voor jullie medewerking aan de DOK studie. Onderzoeksverpleegkundigen **Faiza**, **Lies** en **Esther**, geneeskunde student **Laurence**, en kinderartsen **Gieneke Gonera**, **Lidy Overbeek** en **Jan Uitentuis**, zonder jullie had ik niet zoveel kinderen kunnen includeren! Ook de medewerkers van Certe (huisartsenlaboratorium [**Bert** en **Henk**] en laboratorium voor infectieziekten [**Alewijn Ott** en **Christina**]) en Erasmus Klinische Chemie (**Sascha**) wil ik graag bedanken voor alles wat jullie voor mij gedaan hebben. **Wineke**, bedankt voor het invoeren van de vragenlijsten. **Pieter**, dankzij jouw hulp hebben we ook vragenlijsten online kunnen laten invullen door de deelnemers.

Prof. dr. M.Y. Berger, beste **Marjolein**. Ik had me geen betere promotor kunnen wensen! In de afgelopen jaren heb ik veel van jou mogen leren, vooral tijdens de waardevolle besprekingen. Jouw enthousiasme en kritische vragen hebben mij gemotiveerd om nieuwe dingen uit te proberen en uit te zoeken. Doordat je ruim de tijd nam om mijn manuscripten grondig te bestuderen is mijn proefschrift veel beter geworden. Ik kijk ernaar uit om samen de komende jaren het diagnostisch onderzoek en het diagnostisch proces in de huisartsenpraktijk te verbeteren.

Dr. Y. Lisman-van Leeuwen, beste **Yvonne**. Ondanks dat je twee keer met zwangerschapsverlof was en aan het einde van mijn promotietraject een nieuwe baan hebt gevonden, was je een geweldige dagelijkse begeleidster. Jouw deur stond letterlijk en figuurlijk altijd open. Ik vond het erg fijn dat ik laagdrempelig bij jou kon binnenlopen met mijn vragen, zodat ik weer verder kon. Tijdens congressen, overleggen, pauzes en autoritjes was het met jou altijd gezellig.

Dr. P.F. van Rheenen, beste **Patrick**. Bedankt voor jouw inbreng als specialist op het gebied van kindergastro-enterologie. Ik heb je snelle en kritische commentaren op mijn manuscripten erg gewaardeerd en wil je bedanken dat ik tussendoor altijd even langs kon komen om van gedachten te wisselen over het onderzoek. Door jou zijn mijn figuren een stuk mooier en duidelijker geworden!

De leden van de **beoordelingscommissie**, Prof. dr. G.J. Dinant, Prof. dr. H.J. Verkade, Prof.

dr. P.M.M. Bossuyt, wil ik graag hartelijk bedanken voor hun bereidheid dit proefschrift te lezen en te beoordelen.

Alle medeauteurs wil ik bedanken voor hun bijdrage aan de artikelen en in het bijzonder: Prof. dr. J.C. Escher, dr. A. Kindermann, dr. O.F. Norbruis, Prof. dr. Y.B. de Rijke, dr. B.J. Kollen en dr. J.B. Reitsma. Beste **Hankje**, **Angelika**, **Obbe** en **Yolanda**, bedankt voor jullie bijdrage aan de artikelen over de DOK studie. Beste **Boudewijn**, ik vond het mooi om samen met jou formules uit te pluizen. Ik ben heel dankbaar voor jouw statistische en methodologische hulp en je kritische blik op de manuscripten. Beste **Hans**, bedankt voor de prettige samenwerking. Ik heb veel van jou geleerd over diagnostische meta-analyses.

Naast de bovengenoemden, ben ik dankbaar dat ik mocht leren van andere ervaren onderzoekers: dr. H. Burger, dr. M.H. Blanker, dr. K.H. Groenier. Beste **Huib**, bedankt voor de waardevolle discussies over de methodologische uitdagingen in diagnostisch onderzoek bij patiënten met een lage voorafkans op de ziekte. Hier is een mooie subsidieaanvraag uitgekomen voor een fellowship in Oxford. Beste **Marco**, jij bent van alle markten thuis. Bedankt voor de begeleiding tijdens Yvonne haar zwangerschapsverlof en het lezen en inkorten van manuscripten. Beste **Klaas**, voor ingewikkelde statistische vragen of problemen met STATA kon ik altijd bij jou terecht.

Drie geneeskunde studenten heb ik mogen begeleiden bij hun wetenschappelijke stage. **Geert**, **Jessica** en **Justin**, ik hoop dat jullie net zo veel van mij geleerd hebben als ik van jullie. Justin, ik hoop dat ons artikel over *Dientamoeba Fragilis* snel geaccepteerd wordt. Geert en Jessica, jullie hebben veel dossiers doorgenomen voor het EDDA project. **Anke**, jij hebt hierover een mooi artikel geschreven. Succes met jouw promotie onderzoek!

Karla, **Janny**, **Arnoud**, **Harrie**, en **Betty**, bedankt voor het regelen en ondersteunen bij allerlei praktische dingen omtrent mijn promotieonderzoek.

Aan het begin van dit jaar ben ik begonnen met een nieuw onderzoeksproject bij Prof. dr. S.U. Zuidema en dr. H.J. Luijendijk. Beste **Sytse** en **Dika**, bedankt dat jullie mij de mogelijkheid hebben geboden om mezelf verder te ontwikkelen als onderzoeker.

De collega's binnen de afdeling huisartsgeneeskunde, met name op de **4e verdieping**, wil ik bedanken voor de gezellige en inspirerende werkomgeving. Naast de wetenschappelijke discussies, heb ik ook genoten van de soms te lange koffiepauzes (vaak met taart), lunches, wandelingen, samen sporten (o.a. yoga op de Grote Markt, tennis toernooi en volleybal toernooi), kraambezoeken, diners en borrels. **Chantal**, **Marije**, **Marian** (2x), bedankt voor de gezelligheid tijdens de (externe) cursussen. Gedurende mijn promotieonderzoek heb ik verschillende kamergenoten gehad. **Maarten**, **Thecla**, **Marieke**, **Lieke** en **Jojanneke**, met jullie allemaal heb ik een leuke tijd gehad op onze kleine kamer van twee bij vier meter.

Marian en **Liesbeth**, ik vind het super leuk dat jullie mijn paranimfen willen zijn. Marian, verslapen tijdens een cursus, waarschuwing van politie op de Brooklyn Bridge en het ijskoude

hostel in Harlem waren mooie avonturen. Leuk dat we in november naar de NAPCRG gaan in Colorado Springs! Liesbeth, ik ken je al heel lang en we hebben samen veel geweldige momenten meegemaakt tijdens vakanties, sporten en feestjes. Hoewel het begrip paranimf voor jou nog onbekend was, vind ik het mooi dat je samen met Marian naast mij staat tijdens de verdediging van mijn proefschrift.

Gelukkig is er meer in het leven dan alleen onderzoek doen, **vrienden** bedankt voor de diners, biertjes en wijntjes drinken, weekendjes, fietstochten en skivakanties. **Maria**, **Marleen** en **Liesbeth**, veel dank voor de gezelligheid tijdens onze afspraakjes. Ik hoop dat er nog vele gaan volgen. **Tessa** en **Sanne**, ik heb genoten van de gezellige etentjes waarbij we onze PhD perikelen bespraken en de productieve schrijfweek op Ameland waar ik de inleiding van dit proefschrift heb geschreven. **Marianne**, als jij erbij was, was de AC gang compleet. **Eefje**, fijn dat ik altijd bij jou kon slapen als ik een cursus of congres in Amsterdam had. **Squashmeiden**, ik vond het mooi om elke maandagavond de frustratie er weer uit te slaan. **BWers**, wanneer is het volgende BW weekend?

Beste **familie** van zowel vaders als moeders kant en schoonfamilie, bedankt voor jullie belangstelling tijdens verjaardagen en feestdagen. **Pap**, **Sijbo** & **Natascha**, **Tjitze**, **Ritzo** en **Hendrik**, ik ben erg blij met jullie, opgroeien met vier broers is erg leuk! Ik vind het bijzonder dat ik een aantal jaren later in dezelfde zaal sta als waar ook mijn grote broer Sijbo zijn proefschrift heeft verdedigd. Ritzo, bedankt voor het maken van het mooie logo voor de DOK studie! **Mem**, ik vind het erg jammer dat je deze periode in mijn leven niet meer hebt kunnen meemaken. Dit boekje is voor jou!

Lieve **Maarten**, ik ben erg blij dat ik in de eerste periode van mijn promotie bij jou op de kamer werd geplaatst en dat we nu nog steeds samen gelukkig zijn. Jouw hulp bij dit proefschrift was onmisbaar. Ik heb zin in al onze toekomstige avonturen. Het is geweldig met jou!

DOK
Darm Onderzoek bij Kinderen

LIJST VAN DEELNEMENDE HUISARTSENPRAKTIJKEN

H.P.C. Heijne	Huisartsenpraktijk Heijne en Mulder	Assen
A. Mulder	Huisartsenpraktijk Heijne en Mulder	Assen
M. Sonmez	Huisartsenpraktijk M. Sonmez	Assen
A.H. Tiessen	Huisartsenpraktijk M. Sonmez	Assen
P.G. Tabak	Huisartsenpraktijk P.G. Tabak/ D.J. Nijdam	Assen
D.J. Nijdam	Huisartsenpraktijk P.G. Tabak/ D.J. Nijdam	Assen
F.M.A.M. Verhoeven	Huisartsenpraktijk de Fundatie	Beetsterzwaag
H. Steenbergen	Huisartsenpraktijk Steenbergen	Borger
H.W. Koelemeijer	Huisartsenpraktijk Koelemeijer en Geertjes	Coevorden
B.G.J. Geertjes	Huisartsenpraktijk Koelemeijer en Geertjes	Coevorden
P.A. van Hasselt	Huisartsenpraktijk van Hasselt	Gieten
C. Haverkorn	Huisartsenpraktijk Haverkorn	Emmen
C.D. van Leeuwen	Praktijk van Leeuwen (Quatro)	Emmen
C. Holzhauer	Praktijk Holzhauer (Quattro)	Emmen
P.G.M. van Riswick	Praktijk Holzhauer (Quatro)	Emmen
W.G. Hakman	Huisartsenpraktijk Hakman en Hogervorst (Quattro)	Emmen
S.L. Hogervorst	Huisartsenpraktijk Hakman en Hogervorst (Quattro)	Emmen
J.H. Vaartjes	Huisartsenpraktijk Vaartjes/Rietlanden	Emmen
F.J. Vos	Huisartsenpraktijk Vos en Vos-Noordam	Emmen
M.C. Vos-Noordam	Huisartsenpraktijk Vos en Vos-Noordam	Emmen
H. Groenwold	Huisartsenpraktijk Weerdingerstraat	Emmen
J.W.H. Kocks	Academische Huisartsenpraktijk	Groningen
J.H. Dekker	Academische Huisartsenpraktijk	Groningen
L. Hakvoort	Academische Huisartsenpraktijk	Groningen
E. de Lange	Academische Huisartsenpraktijk	Groningen
J. Talsma	Academische Huisartsenpraktijk	Groningen
P.E. Hermse- de Waard	Academische Huisartsenpraktijk	Groningen
F.A. Said	Academische Huisartsenpraktijk	Groningen
L. AB	Academische Huisartsenpraktijk	Groningen
M. Andeweg	Huisartsenpraktijk Andeweg	Groningen
V. van Vliet	Huisartsenpraktijk Andeweg	Groningen
M.J.J. Heesakker	Huisartsenpraktijk Heesakker en van der Tuin	Groningen
G. van der Tuin	Huisartsenpraktijk Heesakker en van der Tuin	Groningen
P.M. Metternich	Huisartsenpraktijk de Held	Groningen
J.G. Rijkmans	Huisartsenpraktijk Rijkmans	Groningen
J.M. Rijkmans-van der Laan	Huisartsenpraktijk Rijkmans	Groningen
M.G. Hemmen	Zorgpunt Groningen Hoendiep	Groningen
H.J.Th. Hermsen	Zorgpunt Groningen Hoendiep	Groningen
P.M. Salters	Zorgpunt Groningen Hoendiep	Groningen

H.J.H. Wielinga	Zorgpunt Groningen Hoendiep	Groningen
S. Patandin	Huisartsenpraktijk Patandin	Holwierde
G. Geerdink	Huisartsenpraktijk Geerdink	Marum
Y.N. Liemburg-Menzen	Huisartsenpraktijk Nieuw Dordrecht	Nieuw Dordrecht
K.J.T. van der Kamp	Huisartsenpraktijk Reitdiep	Oldehove
J. de Kroon	Apotheekhoudend huisarts Kroon	Onstwedde
W.L. Feis	Huisartsenpraktijk Feis	Oude Pekela
A.C.P. Swierstra	Huisartsenpraktijk Swierstra	Peize
G.B. Groen	Huisartsenpraktijk Groen & Witteveen	Roden
H. Witteveen	Huisartsenpraktijk Groen & Witteveen	Roden
J. Greidanus	Huisartsenpraktijk Statenlaan	Roden
H.H. Hylkema	Huisartsenpraktijk Statenlaan	Roden
M.E.F. van Pelt-Nijman	Huisartsenpraktijk Statenlaan	Roden
C. Peters-Veluthamaningal	Huisartsengroepspraktijk Scheemda	Scheemda
F.B. van Heest	Gezondheidscentrum Schoonoord	Schoonoord
H. Bulk	Huisartsenpraktijk Ter Apel	Ter Apel
H.T. Heslinga	Huisartsenpraktijk Ter Apel	Ter Apel
J.S. Woltjer	Huisartsenpraktijk Usquert	Usquert
S.B. Postma	Huisartsenpraktijk Vlagtwedde	Vlagtwedde
J.D. Berg	Huisartsenpraktijk Wedde	Wedde
H.C. Spijker	Huisartsengroep Oosterbroek - praktijk Noordbroek	Zuidbroek
M.D.A. Brand	Huisartsengroep Oosterbroek - praktijk Zuidbroek	Zuidbroek
J.A. Nikkels	Medisch Centrum Thijinge	Zuidwolde (Dr.)
D. de Groot	Medisch Centrum Thijinge	Zuidwolde (Dr.)
R. Wissink	Medisch Centrum Thijinge	Zuidwolde (Dr.)

CURRICULUM VITAE

Geeske Atje (Gea) Holtman was born on September 10, 1985 in Groningen, the Netherlands. In 2004 she graduated from secondary school at the CSG Augustinus in Groningen and started to study Human Movements Sciences at the University of Groningen. During her master programme she performed a study concerning the validity of the DCD-Daily: a measuring instrument for activities of daily living for children with developmental coordination disorder. This study was supervised by dr. M.M. Schoemaker. In 2009 she obtained her Master of Science in Human Movement Sciences with a specialization in Rehabilitation and Functional Recovery.

From 2010 till 2015 she performed her PhD research at the Department of General Practice at the University Medical Centre Groningen resulting in this thesis. She investigated diagnostic strategies for inflammatory bowel disease in children with chronic gastrointestinal symptoms presenting in primary care under supervision of Prof. dr. M.Y. Berger, general practitioner, dr. Y. Lisman-van Leeuwen, epidemiologist, and dr. P.F. van Rheenen, paediatric gastroenterologist. Notably, she obtained the young investigator award 2015 of the European Society for Paediatric Gastroenterology Hepatology and Nutrition (ESPGHAN) and SHARE PhD top publication award 2015 for the meta-analysis presented in this thesis. During her PhD research she received training as an epidemiologist and obtained her registration as epidemiologist B from the Epidemiological Society in the Netherlands.

Currently, she works as a post-doctoral researcher at the Department of General Practice and Elderly Care Medicine at the University Medical Centre Groningen. She received a grant from the Ter Meulen Grant of the Royal Netherlands Academy of Arts and Sciences for a nine month fellowship at the Nuffield Department of Primary Care Health Sciences at the University of Oxford in 2017. Her research interests are diagnostic and prognostic research in primary care.

LIST OF PUBLICATIONS

PUBLICATIONS IN PEER REVIEWED JOURNALS

International

1. **Holtman GA**, Lisman-van Leeuwen Y, Kollen BJ, Escher JC, Kinderman A, van Rheenen PF, de Rijke YB, Berger MY. Evaluation of point-of-care test calprotectin and lactoferrin for inflammatory bowel disease among children with chronic gastrointestinal symptoms. In press *Family Practice*.
2. **Holtman GA**, Lisman-van Leeuwen Y, Kollen BJ, Escher JC, Kinderman A, Norbruis OF, de Rijke YB, van Rheenen PF, Berger MY. Diagnostic accuracy of fecal calprotectin for pediatric inflammatory bowel disease in primary care; a prospective cohort study. In press *Ann Fam Med*.
3. **Holtman GA**, Lisman-van Leeuwen Y, Reitsma JB, Berger MY. Noninvasive tests for inflammatory bowel disease: a meta-analysis. *Pediatrics*. 2016;137(1):1-11.
4. Heida A, **Holtman GA**, Lisman-van Leeuwen Y, Berger MY, van Rheenen PF. Avoid endoscopy in children with suspected inflammatory bowel disease who have normal calprotectin levels. *J Pediatr Gastroenterol* 2016;62(1):47-49.
5. **Holtman GA**, Lisman-van Leeuwen Y, Kollen BJ, Escher JC, Kinderman A, van Rheenen PF, Berger MY. Challenges in diagnostic accuracy studies in primary care: the fecal calprotectin example. *BMC Fam Pract* 2013;14:179-2296-14-179.
6. Daling AH, **Holtman GA**, Laseur MM, Weinans MA, Lisman-van Leeuwen Y. Observational Research: the Pitfall of Drawing Firm Conclusions About Efficacy. *J Pediatr Gastroenterol Nutr* 2011;53(6):700.

National

7. Burger H, **Holtman GA**. De Receiver Operating Characteristic (ROC)-curve. *Huisarts Wet* 2016; 59(5).
8. van Leeuwen CD, **Holtman GA**, Kranenberg JJ. Kind met buikpijn en rectaal bloedverlies. *Huisarts Wet* 2015; 58(1).
9. Cals JW, Schols AM, van Weert HC, Stevens F, Zeijen CG, **Holtman GA**, Lucassen WAM, Berger MY. Sneltesten in de huisartspraktijk: huidig gebruik en behoefte aan testen in de toekomst. *Ned Tijdschr Geneeskd* 2014;158(0):A8210.
10. **Holtman GA**, Lisman-van Leeuwen Y, Rheenen PF, Berger MY. Kinderen met chronische gastro-intestinale klachten. *Huisarts Wet* 2012; 55(12).

Submitted

11. **Holtman GA**, Lisman-van Leeuwen Y, Kollen BJ, Escher JC, Kinderman A, Norbruis OF, de Rijke YB,van Rheenen PF, Berger MY. Diagnostic test strategies for inflammatory bowel disease in children presenting at primary care level. *Submitted*.
12. **Holtman GA**, Lisman-van Leeuwen Y, Day AS, Fagerberg UL, Henderson P, Leach S, Perminow G, Mack D, van Rheenen PF, Van de Vijver E, Wilson DC, Reitsma JB, Berger MY. Added value of laboratory markers to predict pediatric inflammatory bowel disease: an individual patient data meta-analysis of 1120 patients. *Submitted*.
13. **Holtman GA**, Kranenberg JJ, Blanker MH, Ott A, Lisman-van Leeuwen Y, Berger

MY. *Dientamoeba fragilis* colonization is not associated with gastrointestinal symptoms in children at primary care level. *Submitted*.

14. Lisman-van Leeuwen Y, de Rijke Y, **Holtman GA**, Escher JC, Berger MY. Fecal Calprotectin measurement in children with abdominal pain in primary care increases referral to specialist care. *Submitted*.
15. Van der Hoorn A, van Laar PJ, **Holtman GA**, Westerlaan H. Diagnostic accuracy of Magnetic Resonance Imaging techniques for treatment response evaluation in patients with head and neck tumours, a systematic review and meta-analysis. *Submitted*.
16. Van Dijken BRJ, van Laar PJ, **Holtman GA**, van der Hoorn A. Diagnostic accuracy of Magnetic Resonance Imaging techniques for treatment response evaluation in patients with high-grade glioma, a systematic review and meta-analysis. *Submitted*.

RESEARCH INSTITUTE SHARE

This thesis is published within the Research Institute SHARE (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen.Further information regarding the institute and its research can be obtained from our internetsite: http://www.share.umcg.nl/.

More recent theses can be found in the list below. *((co-) supervisors are between brackets)*

2016

Darvishian M
Real-world influenza vaccine effectiveness; new designs and methods to adjust for confounding and bias
(prof E Hak, prof ER van den Heuvel)

Berm EJJ
Optimizing treatment with psychotropic agents through precision drug therapy; it is not about the mean
(prof B Wilffert, prof E Hak, dr JG Maring)

Fugel H-J
Economics of stratified medicine
(prof MJ Postma, dr M Nuijten, dr K Redekop)

Salavati M
Assessing gross motor function, functional skills, and caregiver assistance in children with cerebral palsy (CP) and cerebral palsy impairment (CPI)
(prof CP van der Schans, prof B Steenbergen, dr A Waninge, dr EAA Rameckers)

Wierike SCM te
'The' pathway towards the elite level in Dutch basketball; a multidimensional and longitudinal study on the development of talented youth basketball players
(prof C Visscher, dr MT Elferink-Gemser)

Jacobs JJWM
General practitioner on an island. How research and innovation help to improve primary care
(prof R Sanderman, prof T van der Molen)

Küpers LK
The first 1000 days and beyond
(prof H Snieder, prof RP Stolk, dr E Corpeleijn)

Kuiper JS
The importance of social relationships in the process of cognitive ageing
(prof RC Oude Voshaar, prof RP Stolk, dr N Smidt)

Göhner, C
Placental particles in pregnancy and preeclampsia
(prof SA Scherjon,prof E Schleußner, dr MM Faas, dr T Plösch)

Vries AJ de
Patellar tendinopathy; causes, consequences and the use of orthoses
(prof RL Diercks, dr I van den Akker-Scheek, dr J Zwerver, dr H van der Worp)

Holland B van
Promotion of sustainable employment; occupational health in the meat processing industry
(prof MF Reneman, prof S Brouwer, dr R Soer, dr MR de Boer)
13.04.2016 EXPAND

Otter TA
Monitoring endurance athletes; a multidisciplinary approach
(prof KAPM Lemmink, prof RL Diercks, dr MS Brink)

Bielderman JH
Active ageing and quality of life; community-dwelling older adults in deprived neighbourhoods
(prof CP van der Schans, dr MHG de Greef, dr GH Schout)

Bijlsma MJ
Age-period-cohort methodology; confounding by birth in cardiovascular pharmacoepidemiology
(prof E Hak, prof S Vansteelandt, dr F Janssen)

Dingemans EAA
Working after retirement; determinants and conzequences of bridge employment
(prof CJIM Henkens, dr ir H van Solinge)

Jonge L de
Data quality and methodology in studies on maternal medication use in relation to congenital anomalies
(prof IM van Langen, prof LTW de Jong-van den Berg, dr MK Bakker)

Vries FM de
Statin treatment in type 2 diabetes patients
(prof E Hak, prof P Denig, prof MJ Postma)

For more 2016 and earlier theses visit our website